

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2004, 09:08:05 ; Search time 57 Seconds
(without alignments)
946.782 Million cell updates/sec

Title: US-09-084-691B-206
Perfect score: 1045
Sequence: 1 MSFLPQKRTKRNTRRP.....CSFSIFLALLSLCTTPASA 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 90 summaries

Database : A_Geneseq_290a04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match Length	ID	Description
1	1045	100.0	191	2	AAR92987 Hepatitis
2	996	95.3	191	2	AAR92972 Hepatitis
3	982	94.0	191	2	AAR92974 Hepatitis
4	982	94.0	191	2	AAR92973 Hepatitis
5	981	93.9	191	2	AAR92953 Hepatitis
6	980	93.8	191	2	AAR92978 Hepatitis
7	978	93.6	191	2	AAR92976 Hepatitis
8	977	93.5	191	2	AAR92977 Hepatitis
9	977	93.5	196	2	AAR74048 Hepatitis
10	977	93.5	196	2	AAR74047 Hepatitis
11	977	93.5	319	2	AAR96546 Hepatitis
12	977	93.5	326	2	AAR922137 HCV-HC59
13	977	93.5	733	2	AAR38278 NMB hepa
14	977	93.5	2894	2	AAR24440 Composite
15	977	93.5	2894	2	AAR70230 Composite
16	977	93.5	3011	2	AAR6995 Hepatitis
17	976	93.4	191	2	AAR44010 Hepatitis
18	976	93.4	191	2	AAR92938 Hepatitis
19	976	93.4	191	2	AAR92941 Hepatitis
20	976	93.4	191	2	AAR92937 Hepatitis
21	976	93.4	191	2	AAR92939 Hepatitis
22	976	93.4	191	2	AAR92940 Hepatitis
23	976	93.4	191	3	AAY94411 Human hep
24	976	93.4	502	2	AAR67591 Hepatitis
25	976	93.4	967	2	AAR79222 pHCV141-e

26	976	93.4	1006	2	AAW12715 HCV genom
27	976	93.4	1648	2	AAR79221 pHCV176-e
28	976	93.4	2984	4	AAE00449 Hepatitis
29	976	93.4	2984	4	AAE00447 Hepatitis
30	976	93.4	2984	4	AAE00442 Hepatitis
31	976	93.4	3011	2	AAR40119 HCV genom
32	976	93.4	3011	2	AAR40120 HCV genom
33	976	93.4	3011	2	AAR79232 HCV sequ
34	976	93.4	3011	2	AAW77397 Hepatitis
35	976	93.4	3011	2	AAW77398 Hepatitis
36	976	93.4	3011	2	AAW98020 Infectiou
37	976	93.4	3011	4	AAE59173 Protein e
38	976	93.4	3011	4	AAE31169 Amino aci
39	976	93.4	3011	5	AAU84597 HCV poly
40	976	93.4	3011	5	AAU799221 Hepatitis
41	976	93.4	3011	5	AAE19888 Hepatitis
42	976	93.4	3011	6	ABP71460 Amino aci
43	976	93.4	3011	6	ABM00339 Hepatitis
44	976	93.4	3012	5	AAU99289 Hepatitis
45	976	93.4	3012	6	ABU61848 HCV H77 c
46	976	93.4	3180	6	ABG73195 MRO-2 vit
47	976	93.4	3208	7	ADD67948 Hepatitis
48	976	93.3	3011	2	AAE67588 Hepatitis
49	974	93.2	191	2	AAE67589 Hepatitis
50	974	93.2	502	2	AAE67589 Hepatitis
51	974	93.2	1031	2	AAE54067 Non-A, no
52	974	93.2	1031	2	AAE98352 5'UTR/COR
53	972	93.0	3011	2	AAU99280 Hepatitis
54	972	93.0	3011	6	AAO26784 Protein d
55	972	93.0	3011	6	ABU61849 HCV-H. 8/
56	971	92.8	385	4	AAE31167 Amino aci
57	970	92.8	191	2	AAE92936 Hepatitis
58	970	92.8	191	2	AAE92932 Hepatitis
59	970	92.8	191	2	AAE92975 Hepatitis
60	970	92.8	513	2	AAE24086 NMB hepa
61	970	92.8	3011	2	AAE22154 NMBV Hut
62	969	92.7	191	2	AAE92949 Hepatitis
63	969	92.7	470	2	AAE34479 Encoded b
64	969	92.7	470	2	AAE34471 Hepatitis
65	969	92.7	729	2	AAE47265 Cyticle p
66	969	92.7	1051	2	AAE54066 Non-A, no
67	969	92.7	1051	2	AAE98361 5'UTR/COR
68	969	92.7	3010	5	ABG32458 Hepatitis
69	969	92.7	3010	5	ABG32459 Hepatitis
70	969	92.7	3010	5	ABG32451 Hepatitis
71	969	92.7	3010	5	ABG32455 Hepatitis
72	969	92.7	3010	5	ABG32457 Hepatitis
73	969	92.7	3010	5	ABG32453 Hepatitis
74	969	92.7	3010	5	ABG32453 Hepatitis
75	969	92.7	3010	5	ABG32461 Hepatitis
76	969	92.7	3010	5	ABG32454 Hepatitis
77	969	92.7	3010	5	ABG32452 Hepatitis
78	969	92.7	3011	5	ABG32456 Hepatitis
79	968	92.6	191	3	AAE82999 Hepatitis
80	968	92.6	2816	2	AAE34009 HCV-1 pol
81	968	92.6	2955	2	AAE14975 Amino aci
82	968	92.6	2955	3	AAE18541 Polynore
83	968	92.6	3011	2	AAE21519 Compiled
84	968	92.6	3011	2	AAE90931 Hepatitis
85	968	92.6	3011	2	AAE34480 HCV poly
86	968	92.6	3011	2	AAE40038 HCV poly
87	968	92.6	3011	5	AAE32049 Hepatitis
88	967	92.5	191	2	AAE37125 Hepatitis
89	967	92.5	191	3	AAE82997 Hepatitis
90	967	92.5	3010	2	AAE30616 Polypept

ALIGNMENTS

RESULT 1
AAR92987 standard; protein; 191 AA.

```
XX AAR92987;
AC HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX DE Hepatitis C virus isolate HK2 core protein.
XX HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX OS Hepatitis C virus.
XX PN WO9605315-A2.
XX PD 22-FEB-1996.
XX PF 15-AUG-1995; 95WO-US010398.
XX PR 15-AUG-1994; 94US-00290665.
XX PS (USSH ) US SEC DEPT HEALTH.
XX P1 Bukh J, Miller RH, Purcell RH;
XX WPI; 1996-139709/14.
XX DR N-PSDB; AAT16661.
XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
XX determine HCV genotype and as vaccines against HCV infection.
XX PS Claim 4; Page 223; 340DP; English.
XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX isolates. Isolated cDNA sequences are used for the prodn. of primers
XX useful for detecting the presence of HCV in a sample, the primers are
XX also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
XX in vaccines for immunising against HCV infection. The proteins may also
XX be used to detect antibodies against HCV in serum, saliva, lymphocytes or
XX other mononuclear cells. The antibodies may be used in the prevention of
XX HCV infection
XX SQ Sequence 191 AA;
Query Match 100.0%; Score 1045; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.4e-90;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RRPPIKARQPGSRHWAQPGYPMPLVGNCGGAGWLLSPRGSRPHWGPNDPRRSRNIG 120
DB 61 RRPPIKARQPGSRHWAQPGYPMPLVGNCGGAGWLLSPRGSRPHWGPNDPRRSRNIG 120
QY 121 KVIDLTCGFADIMGYIPVVGAPVLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTCGFADIMGYIPVVGAPVLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191
RESULT 2
AAR92972
ID AAR92972 standard; protein; 191 AA.
XX
XX AAR92972;
XX 02-OCT-1996 (first entry)
XX Hepatitis C virus isolate Z4 core protein.
```

```
XX HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX OS Hepatitis C virus.
XX PN WO9605315-A2.
XX PD 22-FEB-1996.
XX PF 15-AUG-1995; 95WO-US010398.
XX PR 15-AUG-1994; 94US-00290665.
XX PS (USSH ) US SEC DEPT HEALTH.
XX P1 Bukh J, Miller RH, Purcell RH;
XX WPI; 1996-139709/14.
XX DR N-PSDB; AAT16646.
XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
XX determine HCV genotype and as vaccines against HCV infection.
XX PS Claim 4; Page 210-211; 340DP; English.
XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX isolates. Isolated cDNA sequences are used for the prodn. of primers
XX useful for detecting the presence of HCV in a sample, the primers are
XX also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
XX in vaccines for immunising against HCV infection. The proteins may also
XX be used to detect antibodies against HCV in serum, saliva, lymphocytes or
XX other mononuclear cells. The antibodies may be used in the prevention of
XX HCV infection
XX SQ Sequence 191 AA;
Query Match 95.3%; Score 996; DB 2; Length 191;
Best Local Similarity 94.8%; Pred. No. 9.7e-86;
Matches 181; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RRPPIKARQPGSRHWAQPGYPMPLVGNCGGAGWLLSPRGSRPHWGPNDPRRSRNIG 120
DB 61 RRPPIKARQPGSRHWAQPGYPMPLVGNCGGAGWLLSPRGSRPHWGPNDPRRSRNIG 120
QY 121 KVIDLTCGFADIMGYIPVVGAPVLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTCGFADIMGYIPVVGAPVLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191
RESULT 3
AAR92974
ID AAR92974 standard; protein; 191 AA.
XX
XX AAR92974;
XX 02-OCT-1996 (first entry)
XX Hepatitis C virus isolate Z1 core protein.
XX HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX OS Hepatitis C virus.
```

PN WO9605315-A2.
XX
XX
PD 22-FEB-1996.
XX
PF 15-AUG-1995; 95WO-US010398.
XX
PR 15-AUG-1994; 94US-00290665.
XX
PA (USSH) US SEC DEPT HEALTH.
XX
PI Bukh J, Miller RH, Purcell RH;
XX
PI WPI; 1996-139709/14.
XX
DR N-PSDB; AAT16648.
XX
PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
XX determine HCV genotype and as vaccines against HCV infection.
XX
PS Claim 4; Page 212; 340bp; English.
XX
CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers are
CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
CC in vaccines for immunising against HCV infection. The proteins may also
CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
CC other mononuclear cells. The antibodies may be used in the prevention of
CC HCV infection
XX
SQ Sequence 191 AA;
XX
Query Match 94.0%; Score 982; DB 2; Length 191;
Best Local Similarity 93.7%; Pred. No. 2e-84; Mismatches 5; Indels 0; Gaps 0;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
XX
QY 1 MSTLPKPKQKTKNTNRRPTDVKFPGGGQIVGGVLLPRRGPRLGYRATRKTSERSQPRG 60
DB 1 MSTNPKPKQKTKNTNRRPTDVKFPGGGQIVGGVLLPRRGPRLGYRATRKTSERSQPRG 60
QY 61 RROPIPKARQPOGRHWAQPGYPMPLVGNBCCGAGWLLSPGSRPBMGPNDRRRSRNLG 120
DB 61 RROPIPKARSRSGSWAOPGYPMPLVGNBCCGAGWLLSPGSRPBMGPNDRRRSRNLG 120
QY 121 KYIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KYIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 181 LLSCLTTPASA 191
181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191
XX
RESULT 4
AAR92973
ID AAR92973 standard; protein; 191 AA.
XX
AC AAR92973;
XX
DT 02-OCT-1996 (first entry)
XX
DE Hepatitis C virus isolate z8 core protein.
XX
DE Hepatitis C virus isolate z8 core protein.
XX
KW HCV, E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX
OS Hepatitis C virus.
XX
OS WO9605315-A2.
XX
PN 22-FEB-1996.
XX
PD 15-AUG-1995; 95WO-US010398.
XX
PF 15-AUG-1995; 95WO-US010398.
XX

PR 15-AUG-1994; 94US-00290665.
XX
XX
PA (USSH) US SEC DEPT HEALTH.
XX
PI Bukh J, Miller RH, Purcell RH;
XX
PI WPI; 1996-139709/14.
XX
DR N-PSDB; AAT16647.
XX
PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
XX determine HCV genotype and as vaccines against HCV infection.
XX
PS Claim 4; Page 211-212; 340bp; English.
XX
CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers are
CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
CC in vaccines for immunising against HCV infection. The proteins may also
CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
CC other mononuclear cells. The antibodies may be used in the prevention of
CC HCV infection
XX
SQ Sequence 191 AA;
XX
Query Match 94.0%; Score 982; DB 2; Length 191;
Best Local Similarity 93.7%; Pred. No. 2e-84; Mismatches 5; Indels 0; Gaps 0;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
XX
QY 1 MSTLPKPKQKTKNTNRRPTDVKFPGGGQIVGGVLLPRRGPRLGYRATRKTSERSQPRG 60
DB 1 MSTNPKPKQKTKNTNRRPTDVKFPGGGQIVGGVLLPRRGPRLGYRATRKTSERSQPRG 60
QY 61 RROPIPKARQPOGRHWAQPGYPMPLVGNBCCGAGWLLSPGSRPBMGPNDRRRSRNLG 120
DB 61 RROPIPKARSRSGSWAOPGYPMPLVGNBCCGAGWLLSPGSRPBMGPNDRRRSRNLG 120
QY 121 KYIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KYIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 181 LLSCLTTPASA 191
181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191
XX
RESULT 5
AAR92953
ID AAR92953 standard; protein; 191 AA.
XX
AC AAR92953;
XX
DT 02-OCT-1996 (first entry)
XX
DE Hepatitis C virus isolate HK3 core protein.
XX
DE Hepatitis C virus isolate HK3 core protein.
XX
KW HCV, E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX
OS Hepatitis C virus.
XX
OS WO9605315-A2.
XX
PN 22-FEB-1996.
XX
PD 15-AUG-1995; 95WO-US010398.
XX
PF 15-AUG-1995; 95WO-US010398.
XX
PR 15-AUG-1994; 94US-00290665.
XX
PA (USSH) US SEC DEPT HEALTH.
XX
PI Bukh J, Miller RH, Purcell RH;
XX

DR WPI, 1996-139709/14.
 DR N-PSDB; AAT16627.
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.
 XX
 PS Claim 4; Page 194-195; 340pp; English.
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 CC
 SX Sequence 191 AA;

Query Match 93.9%; Score 981; DB 2; Length 191;
 Best Local Similarity 93.7%; Pred. No. 2.5e-84;
 Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKNTNRRPTDVKFPGGQIVGGVYLLPRRGRLGVRATRKTSERSQPRG 60
 DB 1 MSTNPKPKRKNTNRRPTDVKFPGGQIVGGVYLLPRRGRLGVRATRKTSERSQPRG 60
 QY 61 RROPIPKARQPGRRMAQPGYPMPLYGNEGCGMAGMLSPGSRPHWGPNDRSRRLG 120
 DB 61 RROPIPKARQPGRRMAQPGYPMPLYGNEGCGMAGMLSPGSRPHWGPNDRSRRLG 120
 QY 121 KYIDLTGCFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGYNVATGNLPGCSFSIFLLA 180
 DB 121 KYIDLTGCFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGYNVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 6
 AAR92978
 ID AAR92978 standard; protein; 191 AA.
 XX
 AC AAR92978;
 XX
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate DK13 core protein.
 XX
 KM HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KM hepatitis.
 XX
 OS Hepatitis C virus.
 XX
 PN MO9605315-A2.
 XX
 PD 22-FEB-1996.
 XX
 PF 15-AUG-1995; 95WO-US010398.
 XX
 PR 15-AUG-1994; 94US-00290665.
 XX
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PI Bukh J, Miller RH, Purcell RH;
 XX
 DR WPI, 1996-139709/14.
 DR N-PSDB; AAT16652.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.
 XX

PS Claim 4; Page 215-216; 340pp; English.
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 CC
 SX Sequence 191 AA;

Query Match 93.8%; Score 980; DB 2; Length 191;
 Best Local Similarity 93.7%; Pred. No. 3.1e-84;
 Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKNTNRRPTDVKFPGGQIVGGVYLLPRRGRLGVRATRKTSERSQPRG 60
 DB 1 MSTNPKPKRKNTNRRPTDVKFPGGQIVGGVYLLPRRGRLGVRATRKTSERSQPRG 60
 QY 61 RROPIPKARQPGRRMAQPGYPMPLYGNEGCGMAGMLSPGSRPHWGPNDRSRRLG 120
 DB 61 RROPIPKARQPGRRMAQPGYPMPLYGNEGCGMAGMLSPGSRPHWGPNDRSRRLG 120
 QY 121 KYIDLTGCFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGYNVATGNLPGCSFSIFLLA 180
 DB 121 KYIDLTGCFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGYNVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 7
 AAR92976
 ID AAR92976 standard; protein; 191 AA.
 XX
 AC AAR92976;
 XX
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate Z6 core protein.
 XX
 KM HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KM hepatitis.
 XX
 OS Hepatitis C virus.
 XX
 PN MO9605315-A2.
 XX
 PD 22-FEB-1996.
 XX
 PF 15-AUG-1995; 95WO-US010398.
 XX
 PR 15-AUG-1994; 94US-00290665.
 XX
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PI Bukh J, Miller RH, Purcell RH;
 XX
 DR WPI, 1996-139709/14.
 DR N-PSDB; AAT16650.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.
 XX
 PS Claim 4; Page 214; 340pp; English.
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used

CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection

XX Sequence 191 AA;

Query Match 93.6%; Score 978; DB 2; Length 191;
 Best Local Similarity 93.2%; Pred. No. 4.8e-84;
 Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60
 DB 1 MSTNPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60
 QY 61 RROPITKARQPOGRHWAQPGYPMWPLYGNECCGAGWLLSPRGSRPWGPNDPRRRSRNLG 120
 DB 61 RROPITKARQPOGRHWAQPGYPMWPLYGNECCGAGWLLSPRGSRPWGPNDPRRRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 8
 AAR92977
 ID AAR92977 standard; protein; 191 AA.

XX AAR92977;

XX 02-OCT-1996 (first entry)

XX Hepatitis C virus isolate Z7 core protein.

XX HCV, E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX hepatitis.

XX Hepatitis C virus.

XX MO9605315-A2.

XX 22-FEB-1996.

XX 15-AUG-1995; 95WO-US010398.

XX 15-AUG-1994; 94US-00290665.

XX (USSH) US SEC DEPT HEALTH.

XX Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

XX N-PSDB; AAT16651.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to

XX determine HCV genotype and as vaccines against HCV infection.

XX Claim 4; Page 214-215; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV

XX isolates. Isolated cDNA sequences are used for the prodn. of primers

XX useful for detecting the presence of HCV in a sample, the primers are

XX also useful for HCV genotyping. Proteins encoded by the cDNAs can be used

XX in vaccines for immunising against HCV infection. The proteins may also

XX be used to detect antibodies against HCV in serum, saliva, lymphocytes or

XX other mononuclear cells. The antibodies may be used in the prevention of

Query Match 93.5%; Score 977; DB 2; Length 191;
 Best Local Similarity 93.2%; Pred. No. 6e-84;
 Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60
 DB 1 MSTNPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60
 QY 61 RROPITKARQPOGRHWAQPGYPMWPLYGNECCGAGWLLSPRGSRPWGPNDPRRRSRNLG 120
 DB 61 RROPITKARQPOGRHWAQPGYPMWPLYGNECCGAGWLLSPRGSRPWGPNDPRRRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 9
 AAR74048
 ID AAR74048 standard; protein; 196 AA.

XX AAR74048;

XX 25-MAR-2003 (revised)

XX 26-NOV-1995 (first entry)

XX Synthetic HCV nucleocapsid protein.

XX Hepatitis C virus; vector; expression; prokaryotic cells; detection;

XX diagnosis.

XX Synthetic.

XX MO9511980-A2.

XX 04-MAY-1995.

XX 25-OCT-1994; 94WO-US012166.

XX 25-OCT-1993; 93US-00141917.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Khudyakov Y, Fields HA;

XX WPI; 1995-178872/23.

XX N-PSDB; AAO92071.

XX Vector for expression of a synthetically produced protein coding DNA -

XX pref. encodes the hepatitis C virus nucleocapsid protein which can be

XX used in the detection of HCV antibodies.

XX Disclosure; Page 37; 44pp; English.

XX The sequence is that of a synthetic nucleotide sequence encoding the

XX hepatitis C virus nucleocapsid protein. The gene is positioned in a

XX vector for efficient expression in prokaryotic cells. The expressed

XX protein can be used in tests for the detection of antibodies specific for

XX the HCV C protein. See also AAR74047. (Updated on 25-MAR-2003 to correct

XX PN field.)

XX Sequence 196 AA;

Query Match 93.5%; Score 977; DB 2; Length 196;
 Best Local Similarity 92.1%; Pred. No. 6.1e-84;
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60

```

Db 1 MSTITPKQKTKNTNRPRDQVFPGGGQIVGGVYLPRRGRPLGVRATRKTSERSQPRG 60
Qy 61 RROPPIKAAQPOGRHMAQPGYPMPLVGNBCCGAGMLTSPRGRPRHWGPNDRPRRSRNG 120
Db 61 RROPPIKAAQPOGRHMAQPGYPMPLVGNBCCGAGMLTSPRGRPRHWGPNDRPRRSRNG 120
Qy 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
Qy 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 10
AAR74047
ID AAR74047 standard; protein; 196 AA.
XX
XX AAR74047;
XX
XX 25-MAR-2003 (revised)
XX 26-NOV-1995 (first entry)
XX
XX Synthetic HCV nucleocapsid protein.
XX
XX Hepatitis C virus; vector; expression; prokaryotic cells; detection;
XX diagnosis.
XX
XX Synthetic.
XX
XX MO9511980-A2.
XX
XX 04-MAY-1995.
XX
XX 25-OCT-1994; 94MO-US012166.
XX
XX 25-OCT-1993; 93US-00141917.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Khudyakov Y, Fields HA;
XX
XX WPI: 1995-178872/23.
XX N-PSDB; AAO92070.
XX
XX Vector for expression of a synthetically produced protein coding DNA -
XX pref. encodes the hepatitis C virus nucleocapsid protein which can be
XX used in the detection of HCV antibodies.
XX
XX Disclosure; Page 35; 44pp; English.
XX
XX The sequence is that of a synthetic nucleotide sequence encoding the
XX hepatitis C virus nucleocapsid protein. The gene is positioned in a
XX vector for efficient expression in prokaryotic cells. The expressed
XX protein can be used in tests for the detection of antibodies specific for
XX the HCV C protein. See also AAR74048. (Updated on 25-MAR-2003 to correct
XX FN field.)
XX
XX Sequence 196 AA;
XX
XX Query Match 93.5%; Score 977; DB 2; Length 196;
XX Best Local Similarity 92.1%; Pred. No. 6.1e-84;
XX Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
Qy 1 MSTITPKQKTKNTNRPRDQVFPGGGQIVGGVYLPRRGRPLGVRATRKTSERSQPRG 60
Db 1 MSTITPKQKTKNTNRPRDQVFPGGGQIVGGVYLPRRGRPLGVRATRKTSERSQPRG 60
Qy 61 RROPPIKAAQPOGRHMAQPGYPMPLVGNBCCGAGMLTSPRGRPRHWGPNDRPRRSRNG 120
Db 61 RROPPIKAAQPOGRHMAQPGYPMPLVGNBCCGAGMLTSPRGRPRHWGPNDRPRRSRNG 120
Qy 61 RROPPIKAAQPOGRHMAQPGYPMPLVGNBCCGAGMLTSPRGRPRHWGPNDRPRRSRNG 120
Db 61 RROPPIKAAQPOGRHMAQPGYPMPLVGNBCCGAGMLTSPRGRPRHWGPNDRPRRSRNG 120

```

```

Qy 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
Qy 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 11
AAR96546
ID AAR96546 standard; peptide; 319 AA.
XX
XX AAR96546;
XX
XX 10-MAR-1997 (first entry)
XX
XX Hepatitis C virus types 9a(7a) isolates FRI amino acids 1-317.
XX
XX Hepatitis C virus; subtype; polymerase chain reaction; amplification;
XX PCR; primer; probe; antibody; infection.
XX
XX Hepatitis C virus.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 128
XX FT /label= Phe, Ser, Tyr, Cys
XX FT Misc-difference 129
XX FT /label= Gly
XX FT /note= "amino acid in this position is designated X in
XX the specification, but codon usage shows that the only
XX possible amino acid at this pos. is Gly"
XX FT Misc-difference 308
XX FT /label= Ile, Met
XX FT Misc-difference 310
XX FT /label= Ser, Pro, Thr, Ala
XX
XX MO9613590-A2.
XX
XX 09-MAY-1996.
XX
XX 23-OCT-1995; 95WO-EP004155.
XX
XX 21-OCT-1994; 94EP-00870166.
XX 28-JUN-1995; 95BP-00870076.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Maertens G, Stuyver L;
XX
XX WPI: 1996-251460/25.
XX N-PSDB; AAT77957.
XX
XX Hepatitis C virus poly:nucleic acid unique to unidentified sub-type -
XX used to develop probes and primers for new sub:types and vaccines to
XX prevent and treat infection.
XX
XX Claim 25; Fig 3; 150pp; English.
XX
XX The sequences AAR96526-R96578 represent novel sequences isolated from
XX hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
XX j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-j, 2k, 2l, 3g,
XX 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
XX untranslated region (UR), the core/E1, NS4 or NS5B regions of the genome.
XX This sequence represents amino acids 1-317 from the HCV types 9a and 7a
XX isolates FRI. The new HCV types were isolated from patients with chronic
XX HCV from the Benelux countries, France, Cameroon and Vietnam, because of
XX their aberrant reactivities. The RNA was extracted, cDNA synthesized and
XX PCR amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions
XX were sequenced either directly or partially and used to classify the new
XX viruses into (sub)types based on comparison with known sequences. The
XX sequences were used to generate the peptides AAR96424-R96524. The

```

CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect anti
 CC -HCV antibodies, for HCV typing or to prevent HCV infections

XX Sequence 319 AA;

Query Match 93.5%; Score 977; DB 2; Length 319;

Best Local Similarity 93.7%; Pred. No. 1,1e-83; Mismatches 9; Indels 0; Gaps 0;

Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

Db 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

QY 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

Db 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

QY 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

Db 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

PD 05-MAR-1992.

XX 23-AUG-1991; 91MO-US006037.

XX 25-AUG-1990; 90US-00573643.

XX 21-NOV-1990; 90US-00616369.

XX 21-AUG-1991; 91US-00748564.

XX (NYBL-) NEW YORK BLOO DCENT.

XX (PHAR-) PHARMA.

XX Zebedee S, Inchauspe G, Nasole MS, Prince AM;

XX WPI: 1992-096821/12.

XX N-PADB; AAQ22838.

XX deoxyribonucleic acid sequence encoding non-A, non-B hepatitis virus -

XX obtcd. Hutch C59 subgroup encoding polypeptide(s), useful as vaccines, and

XX immuno reactive ABS for diagnosis of virus.

XX Disclosure, Page 131-133; 225pp; English.

XX One Hutch strain (HCV-H) of NANBV, designated the Hutch C59 isolate (HCV-

XX Hc59) was propagated through passage in animals and the entire viral

XX genome was cloned and sequenced (see AAQ22871). The sequence represented

XX here comprises two amino acids, indicated in the features, which differ

XX from the sequence of AAR22154. The proteins and peptides (see features)

XX and antibodies against them are useful for the prepn. of vaccines and

XX inoculums against NANBV and in immunological assays for detection of

XX viral infection. (Updated on 25-MAR-2003 to correct PR field.) (Updated

XX on 25-MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct

XX OS field.)

XX Sequence 326 AA;

XX Query Match 93.5%; Score 977; DB 2; Length 326;

XX Best Local Similarity 92.1%; Pred. No. 1,1e-83;

XX Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

Db 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

QY 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

Db 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

QY 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

Db 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Query Match 93.5%; Score 977; DB 2; Length 326;

Best Local Similarity 92.1%; Pred. No. 1,1e-83;

Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

Db 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

QY 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

Db 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

QY 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

Db 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

Query Match 93.5%; Score 977; DB 2; Length 326;

Best Local Similarity 92.1%; Pred. No. 1,1e-83;

Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

Db 1 MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

QY 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

Db 61 RROPPIKARPOGRRHMAQPGYPMPLYNBEGCGNAGMLSPRGSPPHMGPNDRRRSRNLG 120

QY 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

Db 121 KVITDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

XX 16-APR-1993.
 PD 91JP-00196175.
 PF 10-APR-1991;
 XX 91JP-00196175.
 XX 12-JUN-1990;
 PR 90JP-00153401.
 XX 08-NOV-1990;
 XX 90JP-00304405.
 PA (NAKA/) NAKAMURA T.
 XX WPI; 1993-199637/25.
 DR N-PSDB; AAQ43888.
 DR
 XX Antigen related to non-A and non-B hepatitis virus - comprises non-
 PT translation region comprising 340 - 341 mols. of nucleotides, non-
 PT translation region comprising 1885 - 2551 mols. of nucleotides including
 PT region 1,149 and, etc.
 XX
 PS Claim 14; Page 32-35; 73pp; Japanese.
 XX
 CC The sequence is that of NANB hepatitis virus HC-OM gene polypeptide P-733
 CC -1. It may be used in a system for detecting NANB hepatitis. This
 CC method is highly specific and sensitive, and can detect NANB hepatitis
 CC virus which could not be detected by conventional methods
 XX
 SQ Sequence 733 AA;

Query Match 93.5%; Score 977; DB 2; Length 733;
 Best Local Similarity 92.1%; Pred. No. 2.7e-83;
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRQKTKNTRRPTDVKFPGGGQIVGGVILLPRGPRLLGYRATKTSERSOPRG 60
 DB 1 MSTLPKPKRQKTKNTRRPTDVKFPGGGQIVGGVILLPRGPRLLGYRATKTSERSOPRG 60
 QY 61 RRQPIPKARQPOGRHWAQPGYPMWLYGNBCCGAGWLLSPRGSRRPHWGPNDPRRSRNIG 120
 DB 61 RRQPIPKARQPOGRHWAQPGYPMWLYGNBCCGAGWLLSPRGSRRPHWGPNDPRRSRNIG 120
 QY 121 KYVDTITCGFADLMGYIPVVGAPLGGVAAALAHGVALEICNYATNGLPFGCSFIFLLA 180
 DB 121 KYVDTITCGFADLMGYIPVVGAPLGGVAAALAHGVALEICNYATNGLPFGCSFIFLLA 180
 QY 181 LUSCLTTPASA 191
 DB 181 LUSCLTTPASA 191
 RE 14
 A 40
 ID AAR24440 standard; protein; 2894 AA.
 AC AAR24440;
 XX
 DT 25-MAR-2003 (revised)
 DT 02-DEC-1992 (first entry)
 XX
 DE Composite HCV HC-J1/CDC/CHI protein.
 XX
 KW Hepatitis C virus; peptides; antibodies; ELISA.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT Peptide /label= 1
 FT Peptide 7..26
 FT Peptide /label= 2
 FT Peptide 13..32
 FT Peptide /label= 3
 FT Peptide 37..56
 FT Peptide /label= 4
 FT Peptide 49..68

FT Peptide /label= 5
 FT Peptide 61..80
 FT Peptide /label= 6
 FT Peptide 73..92
 FT Peptide /label= 7
 FT Peptide 1688..1707
 FT Peptide /label= 8
 FT Peptide 1694..1713
 FT Peptide /label= 9
 FT Peptide 1706..1725
 FT Peptide /label= 10
 FT Peptide 1712..1731
 FT Peptide /label= 11
 FT Peptide 1718..1737
 FT Peptide /label= 12
 FT Peptide 1724..1743
 FT Peptide /label= 13
 FT Peptide 1730..1749
 FT Peptide /label= 14
 FT Peptide 2263..2282
 FT Peptide /label= 15
 FT Peptide 2275..2294
 FT Peptide /label= 16
 FT Peptide 2287..2306
 FT Peptide /label= 17
 FT Peptide 2299..2318
 FT Peptide /label= 18
 FT Peptide 2311..2330
 FT Peptide /label= 19
 PN BP489968-A1.
 XX
 PD 17-JUN-1992.
 XX
 PF 14-DEC-1990; 90EP-00124241.
 PR 14-DEC-1990; 90EP-00124241.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Deleys RJ, Pollet D, Maertens G, Van Heuverswyn H;
 XX
 DR WPI; 1992-201383/25.
 XX
 PT New synthetic peptides for detecting antibodies to hepatitis C virus -
 PT useful in e.g. ELISA assays, and for detection of HCV antigens or as
 PT immunogens.
 PS
 XX Disclosure; Fig 1; 32pp; English.
 XX
 CC RNA viruses frequently exhibit a high rate of spontaneous mutation, thus
 CC a virus is considered to be the same of equiv. to HCV if it exhibits a
 CC global homology of more than 70 percent with the HCV HC-J1/CDC/CHI
 CC composite sequence. The peptide fragments of this DNA sequence indicated
 CC in the features table can immunologically mimic proteins encoded by HCV.
 CC Additional amino acids or chemical groups may be added to either end of the
 CC peptides for the purpose of creating a linker arm for attachment to a
 CC carrier. The peptides can be used for the detection of antibodies
 CC specific for HCV. They may be used in the form of kits, opt. with
 CC reagents such as streptococcal protein A, streptococcal protein G,
 CC avidin or streptavidin. The peptides may also be used as immunogens for
 CC raising antibodies. (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 2894 AA;

Query Match 93.5%; Score 977; DB 2; Length 2894;
 Best Local Similarity 92.1%; Pred. No. 1.2e-82;
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 QY 1 MSTLPKPKRQKTKNTRRPTDVKFPGGGQIVGGVILLPRGPRLLGYRATKTSERSOPRG 60
 DB 1 MSTLPKPKRQKTKNTRRPTDVKFPGGGQIVGGVILLPRGPRLLGYRATKTSERSOPRG 60

QY 61 RROPPIKAROPGRRHMAOPGYPMWPLVGNBCCGAGWLLSPRSGRPHWPNDRRRSRNLG 120
 |||||
 DB 61 RROPPIKARPRBERRTAOPGYPMWPLVGNBCCGAGWLLSPRSGRPHWPNDRRRSRNLG 120
 |||||
 QY 121 KVIDTLTCGFADIMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 DB 121 KVIDTLTCGFADIMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 QY 181 LLSCLTTPASA 191
 |||||
 DB 181 LLSCLTTPASA 191
 |||||

RESULT 15

AA70230
 ID AAR70230 standard; protein; 2894 AA.

XX AAR70230;

XX 5-MAR-2003 (revised)
 DI 07-NOV-1995 (first entry)

XX Composite hepatitis C virus (HC-J1/CDC/CHI).

XX Composite hepatitis C virus; HC-J1/CDC/CHI; HCV; non-A non-B;
 KM synthetic antigens; blood screening.

XX Hepatitis C virus.

XX EP644202-A1.

XX 22-MAR-1995.

XX 14-DEC-1990; 94EP-00108611.

XX 14-DEC-1990; 90EP-00124241.

XX (INNO-) INNOGENETICS NV.

XX Deleys RJ, Pollet D, Maertens G, Van Heuverswyn H;

XX WPI; 1995-116946/16.

XX Synthetic antigens for the detection of hepatitis C virus antibodies -
 PT comprise portions of the HCV peptide sequence, for use in screening blood
 PT and blood products.

XX Disclosure; Fig 1; 51pp; English.

XX AAR70230 is the composite hepatitis C virus (HC-J1/CDC/CHI) protein from
 CC which the synthetic HCV antigens described in AAR70210-R70229 were
 CC derived. These synthetic HCV antigens can be used to screen blood, or blood
 CC products for the presence of HCV, they can also be used in various specific
 CC assays for the detection of HCV antibodies, and antigens, or as
 CC immunogens. (Updated on 25-MAR-2003 to correct PW field.) (Updated on 25-
 CC MAR-2003 to correct PF field.)

XX Sequence 2894 AA;

Query Match 93.5%; Score 977; DB 2; Length 2894;

Best Local Similarity 92.1%; Pred. No. 1.2e-82; Mismatches 9; Indels 0; Gaps 0;

Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPPQKRTKNTNRPTDVKFPGGGQIVGGVLLPRRGRLGVRATRTKTSERSQPRG 60
 |||||
 DB 1 MSTLPPQKRTKNTNRPTDVKFPGGGQIVGGVLLPRRGRLGVRATRTKTSERSQPRG 60
 |||||
 QY 61 RROPPIKAROPGRRHMAOPGYPMWPLVGNBCCGAGWLLSPRSGRPHWPNDRRRSRNLG 120
 |||||
 DB 61 RROPPIKARPRBERRTAOPGYPMWPLVGNBCCGAGWLLSPRSGRPHWPNDRRRSRNLG 120
 |||||
 QY 121 KVIDTLTCGFADIMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 DB 121 KVIDTLTCGFADIMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 |||||

DB 121 KVIDTLTCGFADIMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 QY 181 LLSCLTTPASA 191
 |||||
 DB 181 LLSCLTTPASA 191
 |||||

RESULT 16

AA66995
 ID AAR66995 standard; protein; 3011 AA.

XX AAR66995;

XX 27-AUG-2003 (revised)

DT 01-AUG-1995 (first entry)

XX Hepatitis C virus gene HC-J1/protein.

XX Hepatitis C virus; HCV gene HC-J1/protein; specific antibodies.

XX Hepatitis C virus.

XX JP06284887-A.

XX 11-OCT-1994.

XX 10-DEC-1993; 93JP-00345753.

XX 10-DEC-1992; 92JP-00360705.

XX (IMMO) IMMUNO JAPAN KK.

XX WPI; 1994-362594/45.

XX N-PSDB; AAQ74770.

XX HCV genes and the corresponding proteins - used in the production of anti-
 PT HCV antibodies and the detection of HCV infection.

XX Claim 11; Page 18-32; 35pp; Japanese.

XX AAQ74770 encodes AAR66995 the HC-J1/protein, the cDNA can be used in the
 CC construction of an expression vector for the transformation of a host
 CC cell. The host cell can then be used in the production of proteins and
 CC peptides, useful in the preparation of monoclonal and polyclonal HCV-
 CC specific antibodies. (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 3011 AA;

Query Match 93.5%; Score 977; DB 2; Length 3011;

Best Local Similarity 92.1%; Pred. No. 1.3e-82; Mismatches 9; Indels 0; Gaps 0;

Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPPQKRTKNTNRPTDVKFPGGGQIVGGVLLPRRGRLGVRATRTKTSERSQPRG 60
 |||||
 DB 1 MSTLPPQKRTKNTNRPTDVKFPGGGQIVGGVLLPRRGRLGVRATRTKTSERSQPRG 60
 |||||
 QY 61 RROPPIKAROPGRRHMAOPGYPMWPLVGNBCCGAGWLLSPRSGRPHWPNDRRRSRNLG 120
 |||||
 DB 61 RROPPIKARPRBERRTAOPGYPMWPLVGNBCCGAGWLLSPRSGRPHWPNDRRRSRNLG 120
 |||||
 QY 121 KVIDTLTCGFADIMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 DB 121 KVIDTLTCGFADIMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 QY 181 LLSCLTTPASA 191
 |||||
 DB 181 LLSCLTTPASA 191
 |||||

RESULT 17

AA44010
 ID AAR44010 standard; protein; 191 AA.

XX

AAR92941
 ID AAR92941 standard; protein; 191 AA.
 AC AAR92941;
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate DR4 core protein.
 DE
 XX
 KM HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 XX
 OS Hepatitis C virus.
 PN MO9605315-A2.
 PD 22-FEB-1996.
 PF 15-AUG-1995; 95MO-US010398.
 PR 15-AUG-1994; 94US-00290665.
 PS (USSH) US SEC DEPT HEALTH.
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI; 1996-139709/14.
 DR N-PSDB; AAT16615.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.
 XX
 PS Claim 4; Page 184-185; 340pp; English.
 XX
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 CC
 SQ Sequence 191 AA:
 Query Match 93.4%; Score 976; DB 2; Length 191;
 Best Local Similarity 92.7%; Pred. No. 7.4e-84;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 QY 1 MSTLPKPKORTKNTNRPTDVKEPGGQIVGGVYLLPRGPRPLGVATRTKTSERSOPRG 60
 DB 1 MSTNPKPKORTKNTNRPPDVKEPGGQIVGGVYLLPRGPRPLGVATRTKTSERSOPRG 60
 QY 61 RRPIPKARPOGRHWAQPGYPMPLVGNCGCGAGWLLSPRGRSPHMGPNDRPRRSRLNG 120
 DB 61 RRPIPKARPEGRHTAAGPGYPMPLVGNCGCGAGWLLSPRGRSPHMGPNDRPRRSRLNG 120
 QY 121 KVITDTTCGFADIMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVITDTTCGFADIMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 20
 AAR92937
 ID AAR92937 standard; protein; 191 AA.
 AC AAR92937;
 DT 02-OCT-1996 (first entry)

XX
 DE Hepatitis C virus isolate US11 core protein.
 DE
 KM HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 XX
 OS Hepatitis C virus.
 PN MO9605315-A2.
 PD 22-FEB-1996.
 PF 15-AUG-1995; 95MO-US010398.
 PR 15-AUG-1994; 94US-00290665.
 PS (USSH) US SEC DEPT HEALTH.
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI; 1996-139709/14.
 DR N-PSDB; AAT16611.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.
 XX
 PS Claim 4; Page 181-182; 340pp; English.
 XX
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 CC
 SQ Sequence 191 AA:
 Query Match 93.4%; Score 976; DB 2; Length 191;
 Best Local Similarity 92.7%; Pred. No. 7.4e-84;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 QY 1 MSTLPKPKORTKNTNRPTDVKEPGGQIVGGVYLLPRGPRPLGVATRTKTSERSOPRG 60
 DB 1 MSTNPKPKORTKNTNRPPDVKEPGGQIVGGVYLLPRGPRPLGVATRTKTSERSOPRG 60
 QY 61 RRPIPKARPOGRHWAQPGYPMPLVGNCGCGAGWLLSPRGRSPHMGPNDRPRRSRLNG 120
 DB 61 RRPIPKARPEGRHTAAGPGYPMPLVGNCGCGAGWLLSPRGRSPHMGPNDRPRRSRLNG 120
 QY 121 KVITDTTCGFADIMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVITDTTCGFADIMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 21
 AAR92939
 ID AAR92939 standard; protein; 191 AA.
 AC AAR92939;
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate SW1 core protein.
 DE
 KM HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 XX

OS Hepatitis C virus.
 XX MO9605315-A2.
 XX 22-FEB-1996.
 XX 15-AUG-1995; 95WO-US010398.
 XX 15-AUG-1994; 94US-00290665.
 XX (USSH) US SEC DEPT HEALTH.
 XX Bukh J, Miller RH, Purcell RH;
 XX WPI: 1996-139709/14.
 XX N-PSDB; AAT116613.
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 XX determine HCV genotype and as vaccines against HCV infection.
 XX Claim 4; Page 183; 340pp; English.
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 XX
 SQ Sequence 191 AA;
 Query Match 93.4%; Score 976; DB 2; Length 191;
 Best Local Similarity 92.7%; Pred. No. 7,4e-84;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 DB 1 MSTNPKPKRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 QY 61 RROPIKARPOGRHWAQPGYPMPLVGNCGMAGWLLSPGSRPFWGPNDRRRSRNLG 120
 DB 61 RROPIKARPEGRRTWAQPGYPMPLVGNCGMAGWLLSPGSRPFWGPNDRRRSRNLG 120
 QY 121 KYVDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNATGNLPGCSFSIFLLA 180
 C 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 22
 AAR92940
 ID AAR92940 standard; protein; 191 AA.
 AC AAR92940;
 DT 02-OCT-1996 (first entry)
 XX Hepatitis C virus isolate S18 core protein.
 DE Hepatitis C virus isolate S18 core protein.
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 XX hepatitis.
 XX Hepatitis C virus.
 OS
 XX
 PN MO9605315-A2.
 XX
 PD 22-FEB-1996.

PF 15-AUG-1995; 95WO-US010398.
 XX 15-AUG-1994; 94US-00290665.
 XX (USSH) US SEC DEPT HEALTH.
 XX Bukh J, Miller RH, Purcell RH;
 XX WPI: 1996-139709/14.
 XX N-PSDB; AAT11614.
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 XX determine HCV genotype and as vaccines against HCV infection.
 XX Claim 4; Page 184; 340pp; English.
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 XX
 SQ Sequence 191 AA;
 Query Match 93.4%; Score 976; DB 2; Length 191;
 Best Local Similarity 92.7%; Pred. No. 7,4e-84;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 DB 1 MSTNPKPKRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 QY 61 RROPIKARPOGRHWAQPGYPMPLVGNCGMAGWLLSPGSRPFWGPNDRRRSRNLG 120
 DB 61 RROPIKARPEGRRTWAQPGYPMPLVGNCGMAGWLLSPGSRPFWGPNDRRRSRNLG 120
 QY 121 KYVDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNATGNLPGCSFSIFLLA 180
 C 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 23
 AAY94411
 ID AAY94411 standard; peptide; 191 AA.
 AC AAY94411;
 DT 11-SEP-2000 (first entry)
 XX Human hepatitis C virus core protein.
 DE Human hepatitis C virus core protein.
 KW Human; hepatitis C virus; HCV; immunogen; antigen; antibody; virucide;
 XX hepatotropic; anti-inflammatory; virus detection; vaccine.
 XX Hepatitis C virus.
 OS
 XX
 PN WO200031130-A1.
 XX
 PD 02-JUN-2000.
 XX
 PF 19-NOV-1999; 99WO-IB001933.
 XX
 PR 20-NOV-1998; 98US-00196155.
 XX
 PA (INMR) BIO MERIEUX.
 XX

PI Dalbon P, Jolivet M, Jolivet-Reynaud C;
 XX WPI; 2000-411934/35.
 DR
 XX
 PT Polypeptides that bind to anti-hepatitis C virus antibodies, useful for
 PT diagnosing and preventing hepatitis C infections.
 XX
 PS Disclosure; Fig 1; 50pp; English.
 XX
 CC The present sequence is the core protein of the human hepatitis C virus
 CC (HCV). Three long polypeptide fragments of about 40 amino acids in the
 CC 120 N-terminal amino acids of this sequence were synthesised and their
 CC reactivity towards HCV-positive sera was evaluated in an ELISA test. The
 CC reactivity of the first peptide, S42G, which extends from serine 2 up to
 CC glycine 45, was substantially greater than the reactivity of the other
 CC peptides. Peptide S42G represents an immunodominant region containing
 CC conformational type epitopes and linear type epitopes. S42G manifests an
 CC immunoreactivity with all the sera of individuals or blood samples
 CC infected with HCV and which possess antibodies directed against the core
 CC protein. An amino acid of the S42G sequence may be substituted for
 CC homologous amino acids and side chains and peptide bonds may also be
 CC modified. For example, L-amino acids may be replaced by D-amino acids,
 CC amine groups may be acetylated, and so on. The native antigenic sequence
 CC and its antigenic derivatives may be used for detection of human
 CC hepatitis C viruses and for raising antibodies against the virus
 XX
 SQ Sequence 191 AA;
 Query Match 93.4%; Score 976; DB 3; Length 191;
 Best Local Similarity 92.7%; Pred. No. 7.4e-84;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 QY 1 MSTLPKPKQRTKNTNRRPVDVYKPGGQIVGGVYLLPRGPRPLGVRAATKTSRSQPRG 60
 DB 1 MSTNPKPKQRTKNTNRRPQDVYKPGGQIVGGVYLLPRGPRPLGVRAATKTSRSQPRG 60
 QY 61 RRQPIPKARPOGRHWAQPGYPMPLYGNEGCGWAGMLSPRSGRPMPGNDPRRSRNIG 120
 DB 61 RRQPIPKARREBGRTHWAQPGYPMPLYGNEGCGWAGMLSPRSGRPMPGNDPRRSRNIG 120
 QY 121 KVITDLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINYATGNLPGCSFSIFLLA 180
 DB 121 KVITDLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 24
 AAR67591
 ID AAR67591 standard; protein; 502 AA.
 XX
 AC AAR67591;
 XX
 DT 07-SEP-1995 (first entry)
 XX
 DE Hepatitis C virus SR037-5' gene product.
 XX
 KM Hepatitis C virus; HCV; non-A non-B; SR037-5'; treatment.
 XX
 OS Hepatitis C virus.
 XX
 PN JP06319563-A.
 XX
 PD 22-NOV-1994.
 XX
 PF 13-MAY-1993; 93JP-00147133.
 XX
 PR 13-MAY-1993; 93JP-00147133.
 XX
 PA (IMMO) IMMUNO JAPAN KK.
 XX

DR WPI; 1995-040318/06.
 DR N-PSDB; AAO79143.
 XX
 PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the treatment
 PT of hepatitis C.
 XX
 PS Claim 21; Page 36; 41pp; Japanese.
 XX
 CC AAO79143 is the hepatitis C virus (HCV) gene SR037-5' cDNA, it encodes
 CC the protein described in AAR67591. Both the cDNA and protein can be used
 CC in the treatment of HCV infection
 XX
 SQ Sequence 502 AA;
 Query Match 93.4%; Score 976; DB 2; Length 502;
 Best Local Similarity 92.1%; Pred. No. 2.2e-83;
 Matches 176; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MSTLPKPKQRTKNTNRRPVDVYKPGGQIVGGVYLLPRGPRPLGVRAATKTSRSQPRG 60
 DB 1 MSTNPKPKQRTKNTNRRPQDVYKPGGQIVGGVYLLPRGPRPLGVRAATKTSRSQPRG 60
 QY 61 RRQPIPKARPOGRHWAQPGYPMPLYGNEGCGWAGMLSPRSGRPMPGNDPRRSRNIG 120
 DB 61 RRQPIPKARREBGRTHWAQPGYPMPLYGNEGCGWAGMLSPRSGRPMPGNDPRRSRNIG 120
 QY 121 KVITDLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINYATGNLPGCSFSIFLLA 180
 DB 121 KVITDLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 25
 AAR79222
 ID AAR79222 standard; protein; 967 AA.
 XX
 AC AAR79222;
 XX
 DT 08-DEC-1995 (first entry)
 XX
 DE PHCV141-encoded sequence.
 XX
 KM PHCV141; amyloid precursor protein; APP; hepatitis C virus; HCV; E1; E2;
 KM fusion protein; HEK-293; vaccine; vector; prc/CMV; protein secretion;
 KM glycosylation.
 XX
 OS Synthetic.
 XX
 PN WO9520664-A2.
 XX
 PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-US001087.
 XX
 PR 28-JAN-1994; 94US-00186281.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Watanabe S, Yamaguchi J, Desai SM, Devare SG;
 DR WPI; 1995-275449/36.
 XX
 PT New mammalian expression systems for HCV proteins - express fusion
 PT proteins comprising amyloid precursor protein and HCV E1 and/or E2
 PT protein.
 XX
 PS Disclosure; Page 59-62; 89pp; English.
 XX
 CC RNA from serum or plasma of a chimpanzee infected with HCV was converted
 CC to cDNA and PCR amplified using primers based on HCV sequences. 7

CC Adjacent HCV DNA fragments were generated which together encoded the HCV
CC sequence given in AAR79232. Fragments from 2 clones, pHCV141 and pHCV150
CC (see AAR79223), were combined to generate pHCV176 (see AAR79221). This
CC was used to construct App-HCV-E1+E2 fusion proteins that were
CC glycosylated and secreted from HEK-293 transfectants

SQ Sequence 967 AA;

Query Match 93.4%; Score 976; DB 2; Length 967;
Best Local Similarity 92.7%; Pred. No. 4.5e-83;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDYKFGGQGVVLLPRRGPRLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKPKRTKNTNRPTDYKFGGQGVVLLPRRGPRLGVRATRTKTSERSQPRG 60
QY 61 RROP1PKAROPGRMAQPGYPMPLXGNECGMAGMLSPRGRSPHMGPNDRRRSRNLG 120
DB 61 RROP1PKARPRRGRMAQPGYPMPLXGNECGMAGMLSPRGRSPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 26
AAM12715
ID AAM12715 standard; protein; 1006 AA.

XX AAM12715;

DT 17-OCT-2003 (revised)
DT 09-MAY-1997 (first entry)

DE HCV genome type 1a(H) amino acid residues 1-1006.

XX HCV; influenza virus; vaccine; fusion protein; immunogen; core protein;
XX NS1; plasmid pRT14129.

OS Hepatitis C virus; genotype 1a(H).

XX Key Location/Qualifiers
FH 2..166
FT Region /label= NS1

FT /note= "amino acids 2-166 are utilised in a novel NS1-
F Core fusion protein"

XX W09701640-A2.

XX 16-JAN-1997.

XX 20-JUN-1996; 96WO-EP002764.

XX 29-JUN-1995; 95GB-00013261.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cabezon Silva T, Momin PM, Garcon NMC;

XX WPI; 1997-100211/09.

XX New vaccines against hepatitis C virus - comprising QS21, 3D-MPL, an oil
PT in water emulsion and a HCV core or envelope protein or immunogenic
PT deriv.

XX Example 1; Page 15-16; 20pp; English.

XX Residues 1-1006 of the hepatitis C virus (HCV) genome type 1a(H) are
CC shown in AAM12715. Residues 2-166 (from the core protein) have been

CC utilised in a novel fusion protein (see also AAM12714) with influenza
CC virus NS1. The fusion protein can be formulated into novel vaccines
CC against HCV. (Updated on 17-OCT-2003 to standardise OS field)

SQ Sequence 1006 AA;

Query Match 93.4%; Score 976; DB 2; Length 1006;
Best Local Similarity 92.7%; Pred. No. 4.7e-83;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDYKFGGQGVVLLPRRGPRLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKPKRTKNTNRPTDYKFGGQGVVLLPRRGPRLGVRATRTKTSERSQPRG 60
QY 61 RROP1PKAROPGRMAQPGYPMPLXGNECGMAGMLSPRGRSPHMGPNDRRRSRNLG 120
DB 61 RROP1PKARPRRGRMAQPGYPMPLXGNECGMAGMLSPRGRSPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 27
AAR79221
ID AAR79221 standard; protein; 1648 AA.

XX AAR79221;

DT 08-DEC-1995 (first entry)

DE pHCV176-encoded sequence.

XX pHCV176; amyloid precursor protein; App; hepatitis C virus; HCV; E1; E2;
XX fusion protein; HBK-293; vaccine; vector; pRC/CMV; protein secretion;
XX glycosylation.

OS Synthetic.

XX W09520664-A2.

XX 03-AUG-1995.

XX 27-JAN-1995; 95WO-US001087.

XX 28-JAN-1994; 94US-00188281.

XX (ABBOT) ABBOTT LAB.

XX Watanabe S, Yamaguchi J, Desai SM, Devare SG;

XX WPI; 1995-275449/36.

XX New mammalian expression systems for HCV proteins - express fusion
PT proteins comprising amyloid precursor protein and HCV E1 and/or E2
PT protein.

XX Disclosure; Page 53-59; 89pp; English.

XX RNA from serum or plasma of a chimpanzee infected with HCV was converted
CC to cDNA and PCR amplified using primers based on HCV sequences. 7

CC Adjacent HCV DNA fragments were generated which together encoded the HCV
CC sequence given in AAR79232. Fragments from 2 clones, pHCV141 (see
CC AAR79222) and pHCV150 (see AAR79223), were combined to generate pHCV176.
CC This was used to construct App-HCV-E1+E2 fusion proteins that were
CC glycosylated and secreted from HEK-293 transfectants

XX Sequence 1648 AA;

Query Match 93.4%; Score 976; DB 2; Length 1648;
 Best Local Similarity 92.7%; Pred. No. 8.1e-83;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPPDVKPPGGQIVGGVYLLPRGPRLGVATRTKTSERSQPRG 60
 DB 1 MSTNPKQRTKNTNRPPDVKPPGGQIVGGVYLLPRGPRLGVATRTKTSERSQPRG 60

QY 61 RRQPIPKARPOGRHMAOPGYPMPLVGNBCCGAGMLLSPRGSRPMPGNDPRRRSRLNG 120
 DB 61 RRQPIPKARPEBRTWAQGYPMPLVGNBCCGAGMLLSPRGSRPMPGNDPRRRSRLNG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALANGVRAIEDGIVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALANGVRAIEDGIVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 28
 AAE00449
 ID AAE00449 standard; protein; 2984 AA.

AC AAE00449;
 XX 19-JUN-2001 (first entry)

DE Hepatitis C virus H77C protein lacking HVR1 from chimpanzee #96A008.
 KM Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
 KM antiviral; gene therapy; envelope 2 protein; E2; immunisation; mutant;
 KM HCV infection; viral replication; passive immunoprophylaxis; muten.
 XX Hepatitis C virus.

FT Key Location/Qualifiers
 FT Region 192..383
 FT /label= E1_protein

FT Misc-difference 253
 FT /note= "Wild type Lys substituted with Asn"
 FT Region 384..746
 FT /label= E2_protein
 FT /note= "E2 protein lacks hypervariable region 1"

FT Misc-difference 487
 FT /note= "Wild type Val substituted with Met"
 FT Misc-difference 1116
 FT /note= "Wild type Arg substituted with His"
 FT Misc-difference 1283
 FT /note= "Wild type Tyr substituted with His"

XX WO200121807-A1.
 XX 29-MAR-2001.
 XX 22-SEP-2000; 2000MO-US025987.
 XX 23-SEP-1999; 99US-0155823P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Forms X, Bukh J, Emerson SU, Purcell RH;
 XX WPI; 2001-266076/27.
 XX N-PSDB; AAD03808.
 XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 PT protein lacking all or part of hypervariable region 1 of envelope
 PT protein, useful as vaccine components for treating or preventing HCV
 PT infections.
 XX Example 5; Page; 80pp; English.

XX The present sequence is Hepatitis C virus (HCV) H77C protein from
 CC chimpanzee #96A008 which lacks the hypervariable region one (HVR1) of HCV
 CC envelope 2 (E2). The HCV E2 protein lacking HVR1 (H77C(HVR1)) DNA is
 CC useful for producing infectious HCV and chimeric HCV viruses which are
 CC are useful for identifying cell lines capable of supporting the
 CC replication of viruses. The infectious HCV and HVR1- chimeric HCV are
 CC used in the production of attenuated or inactivated vaccines which are
 CC useful for treating or preventing HCV in a mammal by immunisation. The
 CC host cells expressing the H77C(HVR1) DNA is useful as an immunogen to
 CC stimulate a protective immune response to HCV. The immunogens are useful
 CC for producing protective antibodies to HCV. The antibodies produced are
 CC used in passive immunoprophylaxis for treatment of diseases caused by HCV
 CC in animals, especially humans. The H77C(HVR1) DNA is also useful in gene
 CC therapy. Note: The present sequence is not found in the specification but
 CC is derived from Hepatitis C virus envelope protein 2 lacking
 CC hypervariable region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in
 CC figure 1

XX Sequence 2984 AA;

Query Match 93.4%; Score 976; DB 4; Length 2984;
 Best Local Similarity 92.7%; Pred. No. 1.6e-82;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPPDVKPPGGQIVGGVYLLPRGPRLGVATRTKTSERSQPRG 60
 DB 1 MSTNPKQRTKNTNRPPDVKPPGGQIVGGVYLLPRGPRLGVATRTKTSERSQPRG 60

QY 61 RRQPIPKARPOGRHMAOPGYPMPLVGNBCCGAGMLLSPRGSRPMPGNDPRRRSRLNG 120
 DB 61 RRQPIPKARPEBRTWAQGYPMPLVGNBCCGAGMLLSPRGSRPMPGNDPRRRSRLNG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALANGVRAIEDGIVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALANGVRAIEDGIVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 29
 AAE00447
 ID AAE00447 standard; protein; 2984 AA.

AC AAE00447;
 XX 19-JUN-2001 (first entry)

DE Hepatitis C virus H77C protein lacking HVR1 region from chimpanzee 1590.
 KM Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
 KM antiviral; gene therapy; envelope 2 protein; E2; immunisation; mutant;
 KM HCV infection; viral replication; passive immunoprophylaxis; muten.
 XX Hepatitis C virus.

FT Key Location/Qualifiers
 FT Region 192..383
 FT /label= E1_protein

FT Misc-difference 588
 FT /note= "Wild type Val substituted with Met"
 FT Misc-difference 1116
 FT /note= "Wild type Leu substituted with His"
 FT Misc-difference 1429
 FT /note= "Wild type Arg substituted with His"
 FT Misc-difference 2848
 FT /note= "Wild type Thr substituted with Met"

FT /note= "Wild type Glu substituted with Asp".
 XX
 XX WO200121807-A1.
 XX
 XX 29-MAR-2001.
 XX
 XX 22-SEP-2000; 2000WO-US025987.
 XX
 XX 23-SEP-1999; 99US-0155823P.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Forms X, Bukh J, Emerson SU, Purcell RH,
 XX
 XX MPI: 2001-266076/27.
 XX N-PSDB; AAD03807.
 XX
 XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 XX protein lacking all or part of hypervariable region 1 of envelope
 XX protein, useful as vaccine components for treating or preventing HCV
 XX infections.
 XX
 XX Example 4; Page; 80pp; English.
 XX
 XX The present sequence is Hepatitis C virus (HCV) H77C protein from
 XX chimpanzee 1590 which lacks the hypervariable region one (HVR1) of HCV
 XX envelope 2 (E2). The HCV E2 protein lacking HVR1 (H77C(HVR1)) DNA is
 XX useful for producing infectious HCV and chimeric HCV viruses which are
 XX are useful for identifying cell lines capable of supporting the
 XX replication of viruses. The infectious HCV and HVR1- chimeric HCV are
 XX used in the production of attenuated or inactivated vaccines which are
 XX useful for treating or preventing HCV in a mammal by immunisation. The
 XX host cells expressing the H77C(HVR1)DNA is useful as an immunogen to
 XX stimulate a protective immune response to HCV. The immunogens are useful
 XX for producing protective antibodies to HCV. The antibodies produced are
 XX used in passive immunoprophylaxis for treatment of diseases caused by HCV
 XX in animals, especially humans. The H77C(HVR1) DNA is also useful in gene
 XX therapy. Note: The present sequence is not found in the specification but
 XX is derived from Hepatitis C virus envelope protein 2 lacking
 XX hypervariable region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in
 XX figure 1
 XX
 XX Sequence 2984 AA;
 XX
 XX Query Match 93.4%; Score 976; DB 4; Length 2984;
 XX Best Local Similarity 92.7%; Pred. No. 1.6e-82;
 XX Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 XX
 XX 1 MSTLPKPKRKTNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 XX DB 1 MSTNPKRQRTKNTNRPPDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 XX
 XX 61 RROPPIKAROPGRHMAOPGYPMPLVGNBGGAGMLSPRGSRRPMGPNDRRRSRNLG 120
 XX DB 61 RROPPIKARARBERGTWAOQGYPMPLVGNBGGAGMLSPRGSRRPMGPNDRRRSRNLG 120
 XX
 XX 121 KVIDTLTCGPRADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 XX DB 121 KVIDTLTCGPRADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 XX
 XX 181 LLSCLTTPASA 191
 XX DB 181 LLSCLTTPASA 191
 XX
 XX RESULT 30
 XX AAE00442
 XX ID AAE00442 standard; protein; 2984 AA.
 XX AC AAE00442;
 XX XX
 XX DT 19-JUN-2001 (first entry)
 XX XX

DE Hepatitis C virus envelope 2 protein lacking hypervariable region 1.
 XX
 XX Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
 XX antiviral; gene therapy; envelope 2 protein; E2; immunisation;
 XX HCV infection; viral replication; passive immunoprophylaxis.
 XX
 XX Hepatitis C virus.
 XX
 XX Key Location/Qualifiers
 XX Region 192..383
 XX Region /label= E1_protein
 XX Region 384..746
 XX FT /label= E2_protein
 XX FT /note= "E2 protein lacks hypervariable region 1"
 XX
 XX WO200121807-A1.
 XX
 XX 29-MAR-2001.
 XX
 XX 22-SEP-2000; 2000WO-US025987.
 XX
 XX 23-SEP-1999; 99US-0155823P.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Forms X, Bukh J, Emerson SU, Purcell RH,
 XX
 XX MPI: 2001-266076/27.
 XX N-PSDB; AAD03778.
 XX
 XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 XX protein lacking all or part of hypervariable region 1 of envelope
 XX protein, useful as vaccine components for treating or preventing HCV
 XX infections.
 XX
 XX Claim 1; Fig 1G-H; 80pp; English.
 XX
 XX The present sequence is infectious hepatitis C virus (HCV) of genotype 1a
 XX clone (H77C(HVR1)) which lacks the hypervariable region one (HVR1) of HCV
 XX envelope 2 (E2). The HCV E2 protein lacking HVR1 DNA is useful for
 XX producing infectious HCV and chimeric HCV viruses which are useful
 XX for identifying cell lines capable of supporting the replication of
 XX viruses. The infectious HCV and HVR1-chimeric HCV are used in the
 XX production of attenuated or inactivated vaccines which are useful for
 XX treating or preventing HCV in a mammal by immunisation. The host cells
 XX expressing the H77C(HVR1) DNA is useful as an immunogen to stimulate a
 XX protective immune response to HCV. The immunogens are useful for
 XX producing protective antibodies to HCV. The antibodies produced are used
 XX in passive immunoprophylaxis for treatment of diseases caused by HCV in
 XX animals, especially humans. The H77C(HVR1) DNA is also useful in gene
 XX therapy
 XX
 XX Sequence 2984 AA;
 XX
 XX Query Match 93.4%; Score 976; DB 4; Length 2984;
 XX Best Local Similarity 92.7%; Pred. No. 1.6e-82;
 XX Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 XX
 XX 1 MSTLPKPKRKTNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 XX DB 1 MSTNPKRQRTKNTNRPPDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 XX
 XX 61 RROPPIKAROPGRHMAOPGYPMPLVGNBGGAGMLSPRGSRRPMGPNDRRRSRNLG 120
 XX DB 61 RROPPIKARARBERGTWAOQGYPMPLVGNBGGAGMLSPRGSRRPMGPNDRRRSRNLG 120
 XX
 XX 121 KVIDTLTCGPRADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 XX DB 121 KVIDTLTCGPRADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 XX
 XX 181 LLSCLTTPASA 191
 XX DB 181 LLSCLTTPASA 191

Wed May 12 12:27:17 2004

us-09-084-691b-206.rag

Page 17

Search completed: May 12, 2004, 09:14:36
Job time : 60 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2004, 09:16:06 ; Search time 41 Seconds
(without alignments)
1293.055 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 1045

Sequence: 1 MSLTPKPKQKTKRNTRRPT.....CSPSIFLALSLCTTPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 90 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1045	100.0	319	12	US-10-651-165-228
2	980	93.8	319	12	US-10-651-165-217
3	977	93.5	319	9	US-09-851-138-42
4	977	93.5	319	12	US-10-651-165-229
5	977	93.5	450	12	US-10-651-165-181
6	977	93.5	2894	9	US-09-941-611-23
7	977	93.5	2894	14	US-10-044-995-23
8	976	93.4	249	15	US-10-365-620-54
9	976	93.4	450	12	US-10-651-165-179
10	976	93.4	450	12	US-10-651-165-180
11	976	93.4	473	15	US-10-365-620-56
12	976	93.4	3011	9	US-09-742-659-4
13	976	93.4	3011	9	US-09-952-572-9
14	976	93.4	3011	9	US-09-929-955-1
15	976	93.4	3011	9	US-09-747-419-20

16	976	93.4	3011	10	US-09-891-894-3	Sequence 3, Appl1
17	976	93.4	3011	12	US-10-189-359-14	Sequence 14, Appl
18	976	93.4	3011	12	US-10-296-734-406	Sequence 406, App
19	976	93.4	3011	13	US-10-104-966-1	Sequence 1, Appl1
20	976	93.4	3011	14	US-10-259-275-20	Sequence 20, Appl1
21	976	93.4	3011	14	US-10-184-150-3	Sequence 3, Appl1
22	976	93.4	3011	15	US-10-328-997-3	Sequence 2, Appl1
23	976	93.4	3012	9	US-09-238-076-2	Sequence 2, Appl1
24	976	93.4	3012	10	US-09-995-937-2	Sequence 2, Appl1
25	976	93.4	3012	10	US-09-917-563-2	Sequence 2, Appl1
26	972	93.0	3011	9	US-09-238-076-20	Sequence 20, Appl1
27	972	93.0	3011	10	US-09-995-937-20	Sequence 20, Appl1
28	972	93.0	3011	10	US-09-917-563-20	Sequence 20, Appl1
29	970	92.8	450	12	US-10-651-165-180	Sequence 180, App
30	969	92.7	450	12	US-10-651-165-189	Sequence 189, App
31	969	92.7	3010	12	US-10-467-000-1	Sequence 1, Appl1
32	968	92.6	319	12	US-10-651-165-219	Sequence 219, App
33	968	92.6	3011	9	US-09-916-359-2	Sequence 2, Appl1
34	967	92.5	191	10	US-09-194-949-3	Sequence 187, Appl1
35	965	92.3	450	12	US-10-651-165-187	Sequence 190, App
36	965	92.3	450	12	US-10-651-165-193	Sequence 193, App
37	964	92.2	190	14	US-10-268-562-1	Sequence 1, Appl1
38	964	92.2	319	9	US-09-851-138-48	Sequence 48, Appl1
39	964	92.2	319	12	US-10-651-165-231	Sequence 231, App
40	963	92.2	190	12	US-10-450-649-7	Sequence 7, Appl1
41	962	92.1	809	9	US-09-973-025-50	Sequence 50, Appl1
42	962	92.1	809	10	US-09-899-303-50	Sequence 50, Appl1
43	962	92.1	809	10	US-09-995-808-50	Sequence 50, Appl1
44	962	92.1	809	10	US-09-995-860-50	Sequence 50, Appl1
45	962	92.1	809	10	US-09-995-791-50	Sequence 50, Appl1
46	961	92.0	191	14	US-10-230-381-53	Sequence 53, Appl1
47	961	92.0	209	14	US-10-230-381-7	Sequence 7, Appl1
48	961	92.0	373	14	US-10-230-381-11	Sequence 11, Appl1
49	958	91.7	191	14	US-10-230-381-54	Sequence 54, Appl1
50	958	91.7	373	14	US-10-230-381-13	Sequence 13, Appl1
51	957	91.6	319	10	US-09-899-046-50	Sequence 50, Appl1
52	957	91.6	319	10	US-09-878-281-50	Sequence 50, Appl1
53	957	91.6	319	12	US-10-651-165-226	Sequence 226, App
54	957	91.6	319	12	US-09-873-224-50	Sequence 50, Appl1
55	957	91.6	450	12	US-10-651-165-188	Sequence 188, App
56	956	91.5	191	14	US-10-230-381-55	Sequence 55, Appl1
57	956	91.5	373	14	US-10-230-381-15	Sequence 15, Appl1
58	956	91.5	450	12	US-10-651-165-194	Sequence 194, App
59	956	91.5	450	12	US-10-651-165-195	Sequence 195, App
60	955	91.4	319	12	US-10-651-165-218	Sequence 218, App
61	954	91.3	3011	14	US-10-232-643-6	Sequence 6, Appl1
62	954	91.3	319	10	US-09-899-046-52	Sequence 52, Appl1
63	951	91.0	319	10	US-09-899-046-54	Sequence 54, Appl1
64	951	91.0	319	10	US-09-878-281-52	Sequence 52, Appl1
65	951	91.0	319	10	US-09-878-281-54	Sequence 54, Appl1
66	951	91.0	319	12	US-09-873-224-52	Sequence 52, Appl1
67	951	91.0	319	12	US-09-873-224-54	Sequence 54, Appl1
68	950	90.9	209	14	US-10-230-381-3	Sequence 3, Appl1
69	950	90.9	2985	14	US-10-259-275-40	Sequence 40, Appl1
70	948	90.7	450	12	US-10-651-165-191	Sequence 191, App
71	948	90.7	450	12	US-10-651-165-192	Sequence 192, App
72	946	90.5	189	12	US-10-450-649-9	Sequence 9, Appl1
73	939	89.9	450	12	US-10-651-165-200	Sequence 200, App
74	939	89.9	2240	14	US-10-226-629A-13	Sequence 13, Appl1
75	937	89.7	304	12	US-10-651-165-204	Sequence 204, App
76	937	89.7	319	10	US-09-899-046-144	Sequence 144, App
77	937	89.7	319	10	US-09-878-281-144	Sequence 144, App
78	937	89.7	319	12	US-09-878-224-144	Sequence 201, App
79	936	89.6	450	12	US-10-651-165-201	Sequence 201, App
80	935	89.5	182	9	US-09-929-955-2	Sequence 2, Appl1
81	935	89.5	182	3	US-10-104-966-2	Sequence 2, Appl1
82	934	89.4	190	12	US-10-450-649-8	Sequence 8, Appl1
83	929	88.9	319	9	US-09-851-138-18	Sequence 18, Appl1
84	929	88.9	319	12	US-10-651-165-206	Sequence 206, App
85	924	88.4	319	9	US-09-851-138-44	Sequence 44, Appl1
86	924	88.4	319	12	US-10-651-165-230	Sequence 230, App
87	923	88.3	319	12	US-10-651-165-202	Sequence 202, Appl1
88	915.5	87.6	235	15	US-10-365-620-58	Sequence 58, Appl1

89 915 87.6 459 15 US-10-365-620-60 Sequence 60, Appl
90 895 85.6 319 12 US-10-651-165-211 Sequence 211, App

ALIGNMENTS

RESULT 1
US-10-651-165-228

Sequence 228, Application US/10651165
Publication No. US20040047877A1
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/10/651,165
PRIOR FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 228
LENGTH: 319
TYPE: PR
ORGANISM: hepatitis C virus
US-10-651-165-228

Query Match 100.0%; Score 1045; DB 12; Length 319;
Best Local Similarity 100.0%; Pred. No. 4.3e-85;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPMDVKFPGGGQIVGGVYLLPRGPRLGVAATKTSERSQPRG 60
DB 1 MSTLPKPKRTKNTNRRPMDVKFPGGGQIVGGVYLLPRGPRLGVAATKTSERSQPRG 60
QY 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGMAGWLLSPRSGRPHMGPNDRPRRSRLNG 120
DB 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGMAGWLLSPRSGRPHMGPNDRPRRSRLNG 120
QY 121 KVIDLITCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINYATGNLPGCSFSIFLLA 180
DB 121 KVIDLITCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 2
US-10-651-165-217

Sequence 217, Application US/10651165
Publication No. US20040047877A1
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/10/651,165
PRIOR FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US/08/974,690C
PRIOR FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6

PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 217
LENGTH: 319
TYPE: PR
ORGANISM: hepatitis C virus
US-10-651-165-217

Query Match 93.8%; Score 980; DB 12; Length 319;
Best Local Similarity 93.7%; Pred. No. 2.7e-79;
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPMDVKFPGGGQIVGGVYLLPRGPRLGVAATKTSERSQPRG 60
DB 1 MSTLPKPKRTKNTNRRPMDVKFPGGGQIVGGVYLLPRGPRLGVAATKTSERSQPRG 60
QY 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGMAGWLLSPRSGRPHMGPNDRPRRSRLNG 120
DB 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGMAGWLLSPRSGRPHMGPNDRPRRSRLNG 120
QY 121 KVIDLITCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINYATGNLPGCSFSIFLLA 180
DB 121 KVIDLITCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 3
US-09-851-138-42

Sequence 42, Application US/09851138
Publication No. US20020183508A1
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEBERT
STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS.004
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 319 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-851-138-42

Query Match 93.5%; Score 977; DB 9; Length 319;
Best Local Similarity 93.7%; Pred. No. 5e-79;
Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
DB 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
QY 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120
DB 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120
QY 121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 4

US-10-651-165-229
Sequence 229, Application US/10651165
Publication No. US20040047877A1
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELAYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US/08/974,690C
PRIOR FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 229
LENGTH: 319
TYPE: PRT
ORGANISM: hepatitis C virus
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (128)..(129)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (308)..(308)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (310)..(310)
OTHER INFORMATION: Xaa is any amino acid
US-10-651-165-229

Query Match 93.5%; Score 977; DB 12; Length 319;
Best Local Similarity 93.7%; Pred. No. 5e-79;
Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
DB 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
QY 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120
DB 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120

DB 61 RROPPIKARQPTGRSNGQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120
QY 121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 5

US-10-651-165-181
Sequence 181, Application US/10651165
Publication No. US20040047877A1
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELAYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US/08/974,690C
PRIOR FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 181
LENGTH: 450
TYPE: PRT
ORGANISM: hepatitis C virus
US-10-651-165-181

Query Match 93.5%; Score 977; DB 12; Length 450;
Best Local Similarity 92.1%; Pred. No. 7.4e-79;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
DB 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
QY 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120
DB 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120
QY 121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 6

US-09-941-611-23
Sequence 23, Application US/09941611
Patent No. US20020106640A1
GENERAL INFORMATION:
APPLICANT: DELAYS, ROBERT J
APPLICANT: POULET, DIRK
APPLICANT: MAERTENS, GEERT
APPLICANT: VAN HEUVERSMUN, HUGO
APPLICANT: VAN HEUVERSMUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.

QY 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
DB 1 MSTLPKPKRQKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRILGVATRTKTSRSQPRG 60
QY 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120
DB 61 RROPPIKARQPOGRHWAQPGYPWPLYGNBCCGAGWLLSPRGSRPWMPNDPRRSNLTG 120

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164100
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 93.5%; Score 977; DB 9; Length 2894;
Best Local Similarity 92.1%; Pred. No. 5.8e-78;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRAIRKTSERSQPRG 60
DB 1 MSTLPKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRAIRKTSERSQPRG 60

QY 61 RROPPIKAROPQGRHMAQPGYPMPLVGNBCCGMAWLLSPRGSPPHMGPNDRRRSRNLG 120
DB 61 RROPPIKAROPQGRHMAQPGYPMPLVGNBCCGMAWLLSPRGSPPHMGPNDRRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGIVYATGNLPGCCFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGIVYATGNLPGCCFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 7
US-10-044-995-23
Sequence 23, Application US/10044995
Publication No. US20030049685A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
FOLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSMUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164100
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 93.5%; Score 977; DB 14; Length 2894;
Best Local Similarity 92.1%; Pred. No. 5.8e-78;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRAIRKTSERSQPRG 60
DB 1 MSTLPKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRAIRKTSERSQPRG 60

QY 61 RROPPIKAROPQGRHMAQPGYPMPLVGNBCCGMAWLLSPRGSPPHMGPNDRRRSRNLG 120
DB 61 RROPPIKAROPQGRHMAQPGYPMPLVGNBCCGMAWLLSPRGSPPHMGPNDRRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGIVYATGNLPGCCFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGIVYATGNLPGCCFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 8
US-10-365-620-54
Sequence 54, Application US/10365620
Publication No. US20040001853A1
GENERAL INFORMATION:
APPLICANT: George, Rajan
APPLICANT: Tytrell, Lotne
APPLICANT: No. US20040001853A1jaim, Antoinette

```

? TITLE OF INVENTION: Chimeric Antigens for Eliciting An Immune Response
? FILE REFERENCE: 656.0016
? CURRENT APPLICATION NUMBER: US/10/365,620
? CURRENT FILING DATE: 2003-02-13
? PRIOR APPLICATION NUMBER: US60/423,578
? PRIOR FILING DATE: 2003-11-05
? PRIOR APPLICATION NUMBER: 60/390,564
? PRIOR FILING DATE: 2002-06-20
? NUMBER OF SEQ ID NOS: 76
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 54
? LENGTH: 249
? TYPE: PR1
? ORGANISM: ORF of HCV Core Protein
US-10-365-620-54

```

	Query March Best Local Similarity Batches	93.4%; 92.7%; 177;	Score 976; Pred. No. 4,7e-79; 5;	DB 15; 79; Conservative	Length 249; Indels 0; Gaps 0;
Q1	1	MSTLPKPKORTKRNTRRPDVKFPFGGQIVGGVYLPRGRRLGVRAIRKTSERSQPRG	60		
Db	31	MSTMPKPKORTKRNTRRPDVKFPFGGQIVGGVYLPRGRRLGVRAIRKTSERSQPRG	90		
QY	61	RRORIIPKROPOGRHNWAGPYWPIPLYNEGCGWAGMLSPGSRPHWGNDPRRRSRNLG	120		
Db	91	RROPIPKRRREGKRWAPGPIWPIPLYNEGCGWAGMLSPGSRSPHGWPTDRRRSRNLG	150		
QY	121	KVIDTLTFCFADLNGYIIPVAGAPLPGVAAALAHGVAIRLEDGINVATGNLPQGSFSIFLLA	180		
Db	151	KVIDTLTFCGFADLNGYIIPVAGAPLPGGAARALAHGVAIRLEDGVNATGNLPQGSFSIFLLA	210		
QY	181	LLSCLTTPASA	191		
Db	211	LLSCLTTPASA	221		

```

RESULT 9
US-10-651-165-179
: Sequence 179, Application US/10651165
: Publication No. US20040047877A1
: GENERAL INFORMATION:
: APPLICANT: LEROUX-ROELS, Geert
: APPLICANT: DELEYS, Robert
: APPLICANT: MAERTENS, Geert
: TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
: TITLE OF INVENTION: VIRUS
: PRIORITY REFERENCE: 2551-94
: CURRENT APPLICATION NUMBER: US/10/651,165
: CURRENT FILING DATE: 2003-09-02
: PRIOR APPLICATION NUMBER: US/08/974,690C
: PRIOR FILING DATE: 1997-11-19
: PRIOR APPLICATION NUMBER: PCT/BP94/03555
: PRIOR FILING DATE: 1994-10-28
: PRIOR APPLICATION NUMBER: EP 93402718.6
: PRIOR FILING DATE: 1993-11-04
: NUMBER OF SEQ ID NOS: 286
: SOFTWARE: Patentin version 3.1
: SEQ ID NO 179
: LENGTH: 450
: TYPE: PRT
: ORGANISM: hepatitis C virus
US-10-651-165-179

```

	Query Match	Similarity	Score	ID	Length
Best Local Match	93.4%	92.7%	976	DB 12	450
Matches	177	Conservative	5	Mismatches	9
				Indels	0
				Gaps	0

Db 61 RROPFKARREDEGRWMAQPGYFWPLVYGNBGCMMAGMLTSPGSRSPMGCTPDPRRRSRNLG 120
QY 121 KYIPLTLTGFPADLMKXIPVNGAPRLGVAALAHGVRALPDGINVAQNTPGCSFSIFLLA 180
Db 121 KYIDTLTGFPADLMKXIPVNGAPRLGVAARALAHGVRALPDGINVAQNTPGCSFSIFLLA 180
QY 181 LLSCITTPASA 191
Db 181 LLSCITTPASA 191

```

RESULT 10
US-10-651-165-180
/ Sequence 180. Application US/10651165
/ Publication No. US20040047877A1
/ GENERAL INFORMATION:
/ APPLICANT: LEROUX-ROELS, Geert
/ APPLICANT: DELEYS, Robert
/ APPLICANT: MAERTENS, Geert
/ TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
/ TITLE OF INVENTION: VIRUS
/ FILE REFERENCE: 2551-94
/ CURRENT APPLICATION NUMBER: US/10/651,165
/ CURRENT FILING DATE: 2003-09-02
/ PRIOR APPLICATION NUMBER: US/08/974,690C
/ PRIOR FILING DATE: 1997-11-19
/ PRIOR APPLICATION NUMBER: PCT/BP94/03555
/ PRIOR FILING DATE: 1994-10-28
/ PRIOR APPLICATION NUMBER: EP 93402718.6
/ PRIOR FILING DATE: 1993-11-04
/ NUMBER OF SEQ ID NOS: 286
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 180
/ LENGTH: 450
/ TYPE: prt
/ ORGANISM: hepatitis C virus
US-10-651-165-180

```

```

Query Match Similarity      93.4%; Score 976; DB 12; Length 450;
Best Local Similarity      92.7%; Pred. No. 9, 1e-79;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0

Cy      1  MSTLEPKPQKTKRNTNRREPTDVKPEGGQIVGVYLLPRRGRLGVARTRKTSSENSQPRG 60
Db      1  MSTNPKPQKTKRNTNRREPTDVKFPGGQIVGVYLLPRRGRLGVARTRKTSSENSQPRG 60

Cy      61  RRQPIPKARQPGGRHWAOPGYIPWPLYGNEGCGMAGMLSPRCSRPHMGVNDPRRRSRNG 120
Db      61  RRQPIPKARREGRHWAOPGYIPWPLYGNEGCGMAGMLSPRCSRPHMGVNDPRRRSRNG 120

Cy      121  KVITLITGCFADLTGKGIYIVVGAPLGIVAAALAHGVAIRBGINVATGNLPGCSFSIFLLA 180
Db      121  KVITLITGCFADLTGKGIYIVVGAPLGGAARALAHGVAIRBGINVATGNLPGCSFSIFLLA 180

Cy      181  LLSCLTTPASA 191
Db      181  LLSCLTVPASA 191

RESULT 11
US-10-365-620-56
; Sequence 56, Application US/10365620
; Publication No. US20040001853A1
; GENERAL INFORMATION:
; APPLICANT: George, Rajan
; APPLICANT: Tyrrell, Lorne
; APPLICANT: No. US20040001853A1a1aim, Antioine
; TITLE OF INVENTION: Chimeric Antigens for Eliciting An Immune Response
; FILE REFERENCE: 656.0016
; CURRENT APPLICATION NUMBER: US/10/365,620
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US60/423,578

```

PRIOR FILING DATE: 2003-11-05
PRIOR APPLICATION NUMBER: 60/390,564
PRIOR FILING DATE: 2002-06-20
NUMBER OF SEQ ID NOS: 76
SOFTWARE: Patentin version 3.2
SEQ ID NO 56
LENGTH: 473
TYPE: PRT
ORGANISM: ORF of HCV Core-TBD protein
US-10-365-620-56

Query Match 93.4%; Score 976; DB 15; Length 473;
Best local Similarity 92.7%; Pred. No. 9,5e-79;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPTDVKEPGGGQIVGGVYLLPRGPRLGVAATKTSERSOPRG 60
DB 31 MSTNPKQRTKNTNRPTDVKEPGGGQIVGGVYLLPRGPRLGVAATKTSERSOPRG 90
QY 61 RROPPIKARQPOGRHWAQPGYPMPLVYNEGCGMAGMLSPRGSRPHMGPNDRRRSRNLG 120
DB 91 RROPPIKARQPOGRHWAQPGYPMPLVYNEGCGMAGMLSPRGSRPHMGPNDRRRSRNLG 150
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVAALIEGGINVATGNLPGCSFSIFLLA 180
DB 151 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVAALIEGGINVATGNLPGCSFSIFLLA 210
QY 181 LLSCLTTPASA 191
DB 211 LLSCLTTPASA 221

RESULT 12

US-09-742-659-4
Sequence 4, Application US/09742659
Patent No. US20010034019A1
GENERAL INFORMATION:
APPLICANT: Hong, Zhi
APPLICANT: Butkiewicz, Nancy J.
APPLICANT: Zhong, Weidong
APPLICANT: Ingravallo, Paul
APPLICANT: Wright-Minogue, Jacquelyn
APPLICANT: Lau, Johnson Y.
APPLICANT: Lemon, Stanley M.
TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
FILE REFERENCE: ID01116
CURRENT APPLICATION NUMBER: US/09/742,659
CURRENT FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: US 60/171,469
PRIOR FILING DATE: 1999-12-22
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 3011
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-742-659-4

Query Match 93.4%; Score 976; DB 9; Length 3011;
Best local Similarity 92.7%; Pred. No. 7,5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPTDVKEPGGGQIVGGVYLLPRGPRLGVAATKTSERSOPRG 60
DB 1 MSTNPKQRTKNTNRPTDVKEPGGGQIVGGVYLLPRGPRLGVAATKTSERSOPRG 60
QY 61 RROPPIKARQPOGRHWAQPGYPMPLVYNEGCGMAGMLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARQPOGRHWAQPGYPMPLVYNEGCGMAGMLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVAALIEGGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVAALIEGGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 13

US-09-952-572-9
Sequence 9, Application US/09952572
Patent No. US20020119495A1
GENERAL INFORMATION:
APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
APPLICANT: NAKANO, Eileen
APPLICANT: CLEMENTS, David
APPLICANT: HUMPHREYS, Tom
TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
FILE REFERENCE: HAMBIO100
CURRENT APPLICATION NUMBER: US/09/952,572
CURRENT FILING DATE: 2001-09-13
PRIOR APPLICATION NUMBER: US 60/230,927
PRIOR FILING DATE: 2000-09-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin version 3.0
SEQ ID NO 9
LENGTH: 3011
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-952-572-9

Query Match 93.4%; Score 976; DB 9; Length 3011;
Best local Similarity 92.7%; Pred. No. 7,5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPTDVKEPGGGQIVGGVYLLPRGPRLGVAATKTSERSOPRG 60
DB 1 MSTNPKQRTKNTNRPTDVKEPGGGQIVGGVYLLPRGPRLGVAATKTSERSOPRG 60
QY 61 RROPPIKARQPOGRHWAQPGYPMPLVYNEGCGMAGMLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARQPOGRHWAQPGYPMPLVYNEGCGMAGMLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVAALIEGGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVAALIEGGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 14

US-09-929-955-1
Sequence 1, Application US/09929955
Patent No. US20020136740A1
GENERAL INFORMATION:
APPLICANT: Matti Salberg
APPLICANT: Catharina Hultgren
TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
FILE REFERENCE: TRIPR.23AUS2
CURRENT APPLICATION NUMBER: US/09/929,955
CURRENT FILING DATE: 2001-08-15
PRIOR APPLICATION NUMBER: 09/705,547
PRIOR FILING DATE: 2000-11-03
PRIOR APPLICATION NUMBER: 60/229,175
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/225,767
PRIOR FILING DATE: 2000-08-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3011
TYPE: PRT

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Hepatitis C virus sequence
US-09-929-955-1

Query Match 93.4%; Score 976; DB 9; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
DB 1 MSTNPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARPBGRHTMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 15
US-09-747-419-20
Sequence 20, Application US/09747419
Patent No. US20020155582A1
GENERAL INFORMATION:
APPLICANT: Lemon, Stanley
TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
FILE REFERENCE: 265,0007 0101
CURRENT APPLICATION NUMBER: US/09/747,419
CURRENT FILING DATE: 2000-12-23
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 20
LENGTH: 3011
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Polypeptide
-747-419-20

Query Match 93.4%; Score 976; DB 9; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
DB 1 MSTNPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARPBGRHTMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 16
US-09-891-894-3
Sequence 3, Application US/09891894

Publication No. US20030013081A1
GENERAL INFORMATION:
APPLICANT: Olson, William
APPLICANT: Maddon, Paul
TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITTING HEPATITIS C VIRUS INFI

FILE REFERENCE: 2048/64896/JPM/SHS
CURRENT APPLICATION NUMBER: US/09/891,894
CURRENT FILING DATE: 2001-06-26
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 3011
TYPE: PRT
ORGANISM: hepatitis c virus
US-09-891-894-3

Query Match 93.4%; Score 976; DB 10; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
DB 1 MSTNPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARPBGRHTMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 17
US-10-189-359-14
Sequence 14, Application US/10189359
Publication No. US20040039187A1
GENERAL INFORMATION:
APPLICANT: MARTIN, Annette
APPLICANT: SANGAR, DAVID V.
TITLE OF INVENTION: LEMON, STANLEY M.
FILE REFERENCE: USG:258US
CURRENT APPLICATION NUMBER: US/10/189,359
CURRENT FILING DATE: 2002-07-03
PRIOR FILING DATE: 2002-07-03
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 3011
TYPE: PRT
ORGANISM: Hepatitis C virus
US-10-189-359-14

Query Match 93.4%; Score 976; DB 12; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
DB 1 MSTNPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARPBGRHTMAOPGYPMPLVGNBCCGAGMLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLDEGVVYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 18
US-10-296-734-406
; Sequence 406, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramsdew, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU Pq761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Hepc 1a consensus polypeptide
US-10-296-734-406

Query Match 93.4%; Score 976; DB 12; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
Db 1 MSTNPKPKRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60

QY 61 RROPPIKARPOGRHNAOPGYPMPLVGNBCCGAGWLLSPRGSRPWGPNDPFRSRNLG 120
Db 61 RROPPIKARPEGRTHAOPGYPMPLVGNBCCGAGWLLSPRGSRPWGPNDPFRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLDEGVVYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLDEGVVYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 19
US-10-104-966-1
; Sequence 1, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Macti Salberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Hepatitis C virus sequence
US-10-104-966-1

Query Match 93.4%; Score 976; DB 13; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
Db 1 MSTNPKPKRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60

QY 61 RROPPIKARPOGRHNAOPGYPMPLVGNBCCGAGWLLSPRGSRPWGPNDPFRSRNLG 120
Db 61 RROPPIKARPEGRTHAOPGYPMPLVGNBCCGAGWLLSPRGSRPWGPNDPFRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLDEGVVYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLDEGVVYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 20
US-10-259-275-20
; Sequence 20, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Lemon, Stanley M.
; APPLICANT: Yi, Minkung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCE: 265 0007 0120
; CURRENT APPLICATION NUMBER: US/10/259,275
; PRIOR FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/747,419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325,236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338,123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Polypeptide
US-10-259-275-20

Query Match 93.4%; Score 976; DB 14; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
Db 1 MSTNPKPKRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60

QY 61 RROPPIKARPOGRHNAOPGYPMPLVGNBCCGAGWLLSPRGSRPWGPNDPFRSRNLG 120
Db 61 RROPPIKARPEGRTHAOPGYPMPLVGNBCCGAGWLLSPRGSRPWGPNDPFRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLDEGVVYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLDEGVVYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 21

US-10-184-150-3
Sequence 3, Application US/10184150
Publication No. US20030134297A1
GENERAL INFORMATION:
APPLICANT: Olson, William
APPLICANT: Maddon, Paul
TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITTING HEPATITIS C VIRUS INFECTION
FILE REFERENCE: 2048/64896-A/JPM/MAF/DJK
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: 09/891,894
PRIOR FILING DATE: 2001-06-26
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 3011
TYPE: PRT
ORGANISM: Hepatitis C virus
US-10-184-150-3

Query Match 93.4%; Score 976; DB 14; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
DB 1 MSTNPKRQRTKNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAQPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPTDPRRRSRNLG 120
DB 61 RROPPIKARPEGRHTWAQPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPTDPRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 22

US-10-328-997-3
Sequence 3, Application US/10328997
Publication No. US20030232745A1
GENERAL INFORMATION:
APPLICANT: Olson, William
APPLICANT: Maddon, Paul
TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITTING HEPATITIS C VIRUS INFECTION
FILE REFERENCE: 2048/64896-B/JPM/MAF/DJK
CURRENT FILING DATE: 2002-12-24
PRIOR APPLICATION NUMBER: 09/891,894
PRIOR FILING DATE: 2001-06-26
PRIOR APPLICATION NUMBER: 10/184,150
PRIOR FILING DATE: 2002-06-26
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 3011
TYPE: PRT
ORGANISM: Hepatitis C virus
US-10-328-997-3

Query Match 93.4%; Score 976; DB 15; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

DB 1 MSTNPKRQRTKNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAQPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPTDPRRRSRNLG 120
DB 61 RROPPIKARPEGRHTWAQPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPTDPRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 23

US-09-238-076-2
Sequence 2, Application US/09238076
Patent No. US20020102540A1
GENERAL INFORMATION:
APPLICANT: RICE, CHARLES et al.
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C VIRUS (HCV) AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/238,076
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/034,756
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 6029-4831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3012 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHECAL: NO
FRAGMENT TYPE: N-terminal
US-09-238-076-2

Query Match 93.4%; Score 976; DB 9; Length 3012;
Best Local Similarity 92.7%; Pred. No. 7.5e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
DB 1 MSTNPKRQRTKNTNRRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAQPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPTDPRRRSRNLG 120
DB 61 RROPPIKARPEGRHTWAQPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPTDPRRRSRNLG 120

<i>Qy</i>	122	KVIDTLTCGPAIDMGITPVVGAPLGGVAALAHGRALEDGINATGNLPCCSFSIFLLA	180
<i>Dδ</i>	121	KVIDTLTCTCGPADLMGKITLVGAAPLIGAAAPALAHGVRLVEDGVNATGNLPCCSFSIFLLA	180
<i>Qy</i>	181	LISCLTTTPASA	191
<i>Dδ</i>	181	LISCLTTTPASA	191

RESULT 24
US-09-995

US-09-995-937-2
; Sequence 2, Application US/09995937

Publication No. US20030028010A1
; GENERAL INFORMATION:
;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, I.C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS
STATE: MO
COUNTRY: USA
ZIP: 63105

```

?      COMPUTER READABLE FORM:
?
?          MEDIUM TYPE: Floppy disk
?
?      COMPUTER: IBM PC compatible
?
?      OPERATING SYSTEM: PC-DOS/MS-DOS
?      SOFTWARE: Patent In Release #1.0, Version #1.30
?
?      CURRENT APPLICATION DATA: 9/16/87 12:00 AM

```

[illegible]

Db 181 LLSCLTPASA 191

RESULT 25
US-09-917-563-2

; Sequence 2, Application US/09917563
; Publication No. US20030073080A1

GENERAL INFORMATION:

APPLICANT: RICE, CHARLES et al.
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C

VIRUS (HCV) AND USES TH
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: HOWELL & HAFERKAMP, L. C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS

STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:

```

;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patent in Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:

```

Query Match	93.4%	Score 976;	DB 10;	Length 3012;
Best Local Similarity	92.7%	Pred. No. 7.5e-78;		
Matches 177;	Conservative	5;	Mismatches 9;	Indels 0;
				Gaps 0

QY 181 LLSCLTTPASA 191

RESULT 26

US-09-238-076-20
; Sequence 20, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; US-09-238-076-20

Query Match 93.0%; Score 972; DB 9; Length 3011;
; Best Local Similarity 92.1%; Pred. No. 1,7e-77;
; Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPRPDVYFPGGQIVGVYLLPRGPRLGVAATKTSERSQPRG 60
DB 1 MSTNPKRQRTKNTNRPRPDVYFPGGQIVGVYLLPRGPRLGVAATKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNCGCGMAGMLSPRGSRPHMGPNDRRSRNLG 120
DB 61 RROPPIKARPEGRTHAOPGYPMPLVGNCGCGMAGMLSPRGSRPHMGPNDRRSRNLG 120
QY 121 KVITDLTCGFADLMGYIPVVGAPLGVAAALAHGVARAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVITDLTCGFADLMGYIPVVGAPLGVAAALAHGVARAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 27
US-09-995-937-20
; Sequence 20, Application US/09995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C

VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,937
; FILING DATE: 28-May-1998
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-995-937-20

Query Match 93.0%; Score 972; DB 10; Length 3011;
; Best Local Similarity 92.1%; Pred. No. 1,7e-77;
; Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPRPDVYFPGGQIVGVYLLPRGPRLGVAATKTSERSQPRG 60
DB 1 MSTNPKRQRTKNTNRPRPDVYFPGGQIVGVYLLPRGPRLGVAATKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNCGCGMAGMLSPRGSRPHMGPNDRRSRNLG 120
DB 61 RROPPIKARPEGRTHAOPGYPMPLVGNCGCGMAGMLSPRGSRPHMGPNDRRSRNLG 120
QY 121 KVITDLTCGFADLMGYIPVVGAPLGVAAALAHGVARAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVITDLTCGFADLMGYIPVVGAPLGVAAALAHGVARAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 28
US-09-917-563-20
; Sequence 20, Application US/09917563
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400

CITY: ST. LOUIS
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/917,563
FILING DATE: 27-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/238,076
FILING DATE: 26-JAN-1999
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 6029-4831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-917-563-20

Query Match 93.0%; Score 972; DB 10; Length 3011;
Best Local Similarity 92.1%; Pred. No. 1,7e-77;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRGRPLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRGRPLGVRATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNBGCWAGWLLSPRGSRPMPGNDPRRSRNLG 120
DB 61 RROPPIKARPEGRHMAOPGYPMPLVGNBGCWAGWLLSPRGSRPMPGNDPRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 29
US-10-651-165-190
Sequence 190, Application US/10651165
Publication No. US20040047877A1
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELAYS, Robert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/10/651,165
CURRENT FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US/08/974,690C
PRIOR FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28

PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 190
LENGTH: 450
TYPE: PRT
ORGANISM: hepatitis C virus
US-10-651-165-190

Query Match 92.8%; Score 970; DB 12; Length 450;
Best Local Similarity 92.7%; Pred. No. 3.1e-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRGRPLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRGRPLGVRATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNBGCWAGWLLSPRGSRPMPGNDPRRSRNLG 120
DB 61 RROPPIKARPEGRHMAOPGYPMPLVGNBGCWAGWLLSPRGSRPMPGNDPRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 30
US-10-651-165-189
Sequence 189, Application US/10651165
Publication No. US20040047877A1
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELAYS, Robert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/10/651,165
CURRENT FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US/08/974,690C
PRIOR FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 189
LENGTH: 450
TYPE: PRT
ORGANISM: hepatitis C virus
US-10-651-165-189

Query Match 92.7%; Score 969; DB 12; Length 450;
Best Local Similarity 92.7%; Pred. No. 3.8e-78;
Matches 177; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
QY 1 MSTLPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRGRPLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKPKRKTNTNRRPTDVKFPGGGQIVGGVYLLPRGRPLGVRATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNBGCWAGWLLSPRGSRPMPGNDPRRSRNLG 120
DB 61 RROPPIKARPEGRHMAOPGYPMPLVGNBGCWAGWLLSPRGSRPMPGNDPRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVRALIDGINVATGNLPGCSFSIFLLA 180

Qy	181	LSCLTTPASA	191
Db	181	LSCLTTPASA	191

Search completed: May 12, 2004, 09:21:46
Job time : 43 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2004, 09:13:31 ; Search time 23 Seconds
(without alignments)
428.720 Million cell updates/sec

Title: US-09-084-691B-206
Perfect score: 1045
Sequence: 1 MSFLPKRQKTKRNTNRRT.....CSRSIFLALSLCTTPASA 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/2/1aa/5A COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1045	100.0	191 2 US-08-290-665A-206	Sequence 206, App
2	1045	100.0	191 4 PCT-US95-10398-206	Sequence 206, App
3	1045	100.0	319 4 US-08-635-886C-228	Sequence 228, App
4	1045	100.0	319 4 US-08-974-690C-228	Sequence 228, App
5	996	95.3	191 2 US-08-290-665A-191	Sequence 191, App
6	996	95.3	191 5 PCT-US95-10398-191	Sequence 191, App
7	982	94.0	191 2 US-08-290-665A-192	Sequence 192, App
8	982	94.0	191 2 US-08-290-665A-193	Sequence 193, App
9	982	94.0	191 5 PCT-US95-10398-192	Sequence 192, App
10	982	94.0	191 5 PCT-US95-10398-193	Sequence 193, App
11	981	93.9	191 2 US-08-290-665A-172	Sequence 172, App
12	981	93.9	191 5 PCT-US95-10398-172	Sequence 172, App
13	980	93.8	191 5 US-08-290-665A-197	Sequence 197, App
14	980	93.8	191 5 PCT-US95-10398-197	Sequence 197, App
15	980	93.8	319 4 US-08-635-886C-217	Sequence 217, App
16	980	93.8	319 4 US-08-974-690C-217	Sequence 217, App
17	978	93.6	191 2 US-08-290-665A-195	Sequence 195, App
18	978	93.6	191 5 PCT-US95-10398-195	Sequence 195, App
19	977	93.5	191 2 US-08-290-665A-196	Sequence 196, App
20	977	93.5	191 5 PCT-US95-10398-196	Sequence 196, App
21	977	93.5	319 3 US-08-836-075A-42	Sequence 42, App1
22	977	93.5	319 4 US-08-635-886C-229	Sequence 229, App
23	977	93.5	319 4 US-08-974-690C-229	Sequence 229, App
24	977	93.5	450 4 US-08-635-886C-181	Sequence 181, App
25	977	93.5	450 4 US-08-974-690C-181	Sequence 181, App
26	977	93.5	2894 2 US-08-466-975A-23	Sequence 23, App1
27	977	93.5	2894 2 US-08-391-671A-23	Sequence 23, App1

28	977	93.5	2894 3 US-08-467-902A-23	Sequence 23, App1
29	977	93.5	2894 3 US-09-275-265-23	Sequence 23, App1
30	977	93.5	2894 4 US-09-941-611-23	Sequence 23, App1
31	976	93.4	191 2 US-08-290-665A-156	Sequence 156, App
32	976	93.4	191 2 US-08-290-665A-157	Sequence 157, App
33	976	93.4	191 2 US-08-290-665A-158	Sequence 158, App
34	976	93.4	191 2 US-08-290-665A-159	Sequence 159, App
35	976	93.4	191 2 US-08-290-665A-160	Sequence 160, App
36	976	93.4	191 2 US-08-290-665A-161	Sequence 161, App
37	976	93.4	191 5 PCT-US95-10398-156	Sequence 156, App
38	976	93.4	191 5 PCT-US95-10398-157	Sequence 157, App
39	976	93.4	191 5 PCT-US95-10398-158	Sequence 158, App
40	976	93.4	191 5 PCT-US95-10398-159	Sequence 159, App
41	976	93.4	191 5 PCT-US95-10398-160	Sequence 160, App
42	976	93.4	450 4 US-08-635-886C-179	Sequence 179, App
43	976	93.4	450 4 US-08-635-886C-180	Sequence 180, App
44	976	93.4	450 4 US-08-974-690C-179	Sequence 179, App
45	976	93.4	450 4 US-08-974-690C-180	Sequence 180, App
46	976	93.4	967 1 US-08-188-281B-13	Sequence 13, App1
47	976	93.4	967 5 PCT-US94-07280-13	Sequence 13, App1
48	976	93.4	967 5 PCT-US95-01087-13	Sequence 13, App1
49	976	93.4	1648 1 US-08-188-281B-12	Sequence 12, App1
50	976	93.4	1648 1 PCT-US94-07280-12	Sequence 12, App1
51	976	93.4	1648 5 PCT-US95-01087-12	Sequence 12, App1
52	976	93.4	3011 1 US-08-188-281B-11	Sequence 11, App1
53	976	93.4	3011 1 US-08-453-552-1	Sequence 1, App1
54	976	93.4	3011 1 US-08-453-552-2	Sequence 2, App1
55	976	93.4	3011 2 US-08-710-637-1	Sequence 1, App1
56	976	93.4	3011 2 US-08-710-637-2	Sequence 2, App1
57	976	93.4	3011 4 US-09-014-416-1	Sequence 1, App1
58	976	93.4	3011 4 US-10-104-966-1	Sequence 1, App1
59	976	93.4	3011 5 US-09-952-572-9	Sequence 9, App1
60	976	93.4	3011 5 PCT-US93-00907-1	Sequence 1, App1
61	976	93.4	3011 5 PCT-US93-00907-2	Sequence 2, App1
62	976	93.4	3011 5 PCT-US94-07280-1	Sequence 1, App1
63	976	93.4	3011 5 PCT-US95-01087-1	Sequence 1, App1
64	976	93.4	3012 4 US-08-811-566-2	Sequence 2, App1
65	976	93.4	3012 4 US-09-034-756-2	Sequence 2, App1
66	974	93.2	191 2 US-08-290-665A-174	Sequence 174, App
67	974	93.2	191 5 PCT-US95-10398-174	Sequence 174, App
68	972	93.0	3011 4 US-08-811-566-20	Sequence 20, App1
69	972	93.0	3011 4 US-09-034-756-20	Sequence 20, App1
70	971	92.9	191 2 US-08-290-665A-166	Sequence 166, App
71	971	92.9	191 5 PCT-US95-10398-166	Sequence 166, App
72	971	92.9	340 1 US-08-462-195-2	Sequence 2, App1
73	971	92.9	340 2 US-08-636-883-2	Sequence 2, App1
74	971	92.9	340 3 US-09-127-829-2	Sequence 2, App1
75	970	92.8	191 2 US-08-290-665A-155	Sequence 155, App
76	970	92.8	191 2 US-08-290-665A-171	Sequence 171, App
77	970	92.8	191 2 US-08-290-665A-194	Sequence 194, App
78	970	92.8	191 5 PCT-US95-10398-155	Sequence 155, App
79	970	92.8	191 5 PCT-US95-10398-171	Sequence 171, App
80	970	92.8	191 5 PCT-US95-10398-194	Sequence 194, App
81	970	92.8	450 4 US-08-635-886C-190	Sequence 190, App
82	970	92.8	450 4 US-08-974-690C-190	Sequence 190, App
83	969	92.7	191 2 US-08-290-665A-168	Sequence 168, App
84	969	92.7	191 5 PCT-US95-10398-167	Sequence 167, App
85	969	92.7	191 5 PCT-US95-10398-168	Sequence 168, App
86	969	92.7	191 5 PCT-US95-10398-169	Sequence 169, App
87	969	92.7	450 4 US-08-635-886C-189	Sequence 189, App
88	969	92.7	450 4 US-08-974-690C-189	Sequence 189, App
89	968	92.6	191 4 US-09-763-260-4	Sequence 4, App1
90	968	92.6	319 4 US-08-635-886C-219	Sequence 219, App

ALIGNMENTS

RESULT 1
US-08-290-665A-206
; Sequence 206, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:

APPLICANT: BUKH, J., MILLER, R. H. AND
APPLICANT: PURCELL, R. H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS: 263
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ. ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: HK2
US-08-290-665A-206

Query Match 100.0%; Score 1045; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 7.4e-97;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSTLPRKQRTKNTNRPTDVKFPGGQIVGGVLLPRRGPRLGVATRTKTSERSQPRG 60
1 MSTLPRKQRTKNTNRPTDVKFPGGQIVGGVLLPRRGPRLGVATRTKTSERSQPRG 60
61 RROPPIKAROPQGRHQAQGYPMPLYGNCGGAGWLLSPRGRPHWGPNDRRRSRNLG 120
61 RROPPIKAROPQGRHQAQGYPMPLYGNCGGAGWLLSPRGRPHWGPNDRRRSRNLG 120
121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
181 LLSCLTTPASA 191
181 LLSCLTTPASA 191

RESULT 2
PCT-US95-10398-206
Sequence 206, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R. H. AND
APPLICANT: PURCELL, R. H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS: 263
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ. ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: HK2
PCT-US95-10398-206

Query Match 100.0%; Score 1045; DB 5; Length 191;
Best Local Similarity 100.0%; Pred. No. 7.4e-97;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSTLPRKQRTKNTNRPTDVKFPGGQIVGGVLLPRRGPRLGVATRTKTSERSQPRG 60
1 MSTLPRKQRTKNTNRPTDVKFPGGQIVGGVLLPRRGPRLGVATRTKTSERSQPRG 60
61 RROPPIKAROPQGRHQAQGYPMPLYGNCGGAGWLLSPRGRPHWGPNDRRRSRNLG 120
61 RROPPIKAROPQGRHQAQGYPMPLYGNCGGAGWLLSPRGRPHWGPNDRRRSRNLG 120
121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
121 KVIDTLTCGRADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
181 LLSCLTTPASA 191
181 LLSCLTTPASA 191

RESULT 3
US-08-635-886C-228
Sequence 228, Application US/08635886C
Patent No. 6555114
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert

;; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
;; FILE REFERENCE: VIRUS
;; CURRENT APPLICATION NUMBER: US/08/635,886C
;; PRIOR FILING DATE: 1996-04-25
;; PRIOR APPLICATION NUMBER: PCT/EP94/03555
;; PRIOR FILING DATE: 1994-10-28
;; PRIOR APPLICATION NUMBER: EP 93402718.6
;; PRIOR FILING DATE: 1993-11-04
;; NUMBER OF SEQ ID NOS: 286
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 228
;; LENGTH: 319
;; TYPE: PR1
;; ORGANISM: hepatitis C virus
US-08-635-886C-228

Query Match 100.0%; Score 1045; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.4e-96;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPPQRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLVGATRTKTSERSQPRG 60

DB 1 MSTLPPQRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLVGATRTKTSERSQPRG 60

QY 61 RROPITKARPOGRHMAOPGYPMPLVGNBCCGAGWLLSPRGSRPMPGPNDRRRSRNLG 120

DB 61 RROPITKARPOGRHMAOPGYPMPLVGNBCCGAGWLLSPRGSRPMPGPNDRRRSRNLG 120

QY 121 KVIDITLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

DB 121 KVIDITLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191

DB 61 RROPITKARPOGRHMAOPGYPMPLVGNBCCGAGWLLSPRGSRPMPGPNDRRRSRNLG 120
QY 121 KVIDITLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDITLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 5

US-08-290-665A-191

Sequence 191, Application US/08290665A

Patent No. 5882852

GENERAL INFORMATION:

APPLICANT: BORK, J., MILLER, R.H. AND

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & PINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 191:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: hominids

INDIVIDUAL ISOLATE: Z4

US-08-290-665A-191

Query Match 95.3%; Score 996; DB 2; Length 191;

Best Local Similarity 94.8%; Pred. No. 5.9e-92;

Matches 181; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTLPPQRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLVGATRTKTSERSQPRG 60

DB 1 MSTLPPQRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLVGATRTKTSERSQPRG 60

QY 61 RROPITKARPOGRHMAOPGYPMPLVGNBCCGAGWLLSPRGSRPMPGPNDRRRSRNLG 120

DB 61 RROPITKARPOGRHMAOPGYPMPLVGNBCCGAGWLLSPRGSRPMPGPNDRRRSRNLG 120

QY 121 KVIDITLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

DB 121 KVIDITLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191

DB 181 LLSCLTTPASA 191

Db 121 KVIDLTLCGFADIMGYIPVGAIVGVAAALAHGVAEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 6

PCT-US95-10398-191
Sequence 191, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BURKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 191:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z4
PCT-US95-10398-191

Query Match 95.3%; Score 996; DB 5; Length 191;
Best Local Similarity 94.8%; Pred. No. 5.9e-92;
Matches 181; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTLPRKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
Db 1 MSTNPKRQRTKNTNRPMVDVFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHWAOPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPNDRRRSRNIG 120
Db 61 RROPPIKARPOGRHWAOPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPNDRRRSRNIG 120
QY 121 KVIDLTLCGFADIMGYIPVGAIVGVAAALAHGVAEDGINVATGNLPGCSFSIFLLA 180

Db 121 KVIDLTLCGFADIMGYIPVGAIVGVAAALAHGVAEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 7

US-08-290-665A-192
Sequence 192, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 192:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z8
US-08-290-665A-192

Query Match 94.0%; Score 982; DB 2; Length 191;
Best Local Similarity 93.7%; Pred. No. 1.5e-90;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPRKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
Db 1 MSTNPKRQRTKNTNRPMVDVFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
QY 61 RROPPIKARPOGRHWAOPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPNDRRRSRNIG 120
Db 61 RROPPIKARPOGRHWAOPGYPMPLVGNBCCGAGWLLSPRGRSPHMGPNDRRRSRNIG 120
QY 121 KVIDLTLCGFADIMGYIPVGAIVGVAAALAHGVAEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDLTLCGFADIMGYIPVGAIVGVAAALAHGVAEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191

Db 181 L1SCLTVPASA 191

RESULT 8
US-08-290-665A-193
Sequence 193, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R. H. AND
APPLICANT: PURCELL, R. H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z1
US-08-290-665A-193

Query Match 94.0%; Score 982; DB 2; Length 191;
Best Local Similarity 93.7%; Pred. No. 1.5e-90;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 181 L1SCLTVPASA 191

RESULT 9
PCT-US95-10398-192
Sequence 192, Application PC/TUS9510398

GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R. H. AND
APPLICANT: PURCELL, R. H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 192:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z8
PCT-US95-10398-192

Query Match 94.0%; Score 982; DB 5; Length 191;
Best Local Similarity 93.7%; Pred. No. 1.5e-90;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 181 L1SCLTVPASA 191

RESULT 10
PCT-US95-10398-193
Sequence 193, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUCH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & PINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z1
PCT-US95-10398-193
Query Match 94.0%; Score 982; DB 5; Length 191;
Best Local Similarity 93.7%; Pred. No. 1.5e-90;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKXNTNRPTDVKFGGQIVGGVYLLPRGPRLVGVARTRKTSERSQPRG 60
DB 1 MSTNPKQRTKXNTNRPTDVKFGGQIVGGVYLLPRGPRLVGVARTRKTSERSQPRG 60
QY 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGMAWLLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGMAWLLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 11
US-08-290-665A-172
Sequence 172, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUCH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & PINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 172:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: HK3
US-08-290-665A-172
Query Match 93.9%; Score 981; DB 2; Length 191;
Best Local Similarity 93.7%; Pred. No. 1.9e-90;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKXNTNRPTDVKFGGQIVGGVYLLPRGPRLVGVARTRKTSERSQPRG 60
DB 1 MSTNPKQRTKXNTNRPTDVKFGGQIVGGVYLLPRGPRLVGVARTRKTSERSQPRG 60
QY 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGMAWLLSPRGSRPHMGPNDRRRSRNLG 120
DB 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGMAWLLSPRGSRPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 12
PCT-US95-10398-172

Sequence 172, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCY/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 172:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: HK3
US95-10398-172

Query Match 93.9%; Score 981; DB 5; Length 191;
Best Local Similarity 93.7%; Pred. No. 1.9e-90;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPPQRTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60
DB 1 MSTNPPQRTKNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60

QY 61 RROPITKAROPQGRHNAOPGYPMPLVGNCGCGNAGWLLSPRGRPHMGPNDRRRSRNLG 120
DB 61 RROPITKAROPQGRHNAOPGYPMPLVGNCGCGNAGWLLSPRGRPHMGPNDRRRSRNLG 120

QY 121 KVIDLTCGADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTCGADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 13

US-08-290-665A-197
Sequence 197, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 197:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK13
US-08-290-665A-197

Query Match 93.8%; Score 980; DB 2; Length 191;
Best Local Similarity 93.7%; Pred. No. 2.3e-90;
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPQRTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60
DB 1 MSTNPPQRTKNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLVGVRATKTSERSQPRG 60

QY 61 RROPITKAROPQGRHNAOPGYPMPLVGNCGCGNAGWLLSPRGRPHMGPNDRRRSRNLG 120
DB 61 RROPITKAROPQGRHNAOPGYPMPLVGNCGCGNAGWLLSPRGRPHMGPNDRRRSRNLG 120

QY 121 KVIDLTCGADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTCGADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 14
PCT-US95-10398-197
Sequence 197, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DECODED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 197:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK13
PCT-US95-10398-197

Query Match 93.8%; Score 980; DB 5; Length 191;
Best Local Similarity 93.7%; Pred. No. 2.3e-90;
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPKQRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKQRTKNTNRRPMDVDFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
QY 61 RRQPIPARQPOGRHWAQPGYPMPLVYNGECGAGWLLSPRGSRRPHGPNDRRRSRNLG 120
DB 61 RRQPIPARQPOGRHWAQPGYPMPLVYNGECGAGWLLSPRGSRRPHGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 15
US-08-635-886C-217
Sequence 217, Application US/08635886C
Patent No. 6555114

GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
TITLE OF INVENTION: VIRUS
FILE REFERENCE: 2752-18
CURRENT APPLICATION NUMBER: US/08/635,886C
CURRENT FILING DATE: 1996-04-25
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: Patentin version 3.1
SEQ ID NO 217
LENGTH: 319
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-635-886C-217

Query Match 93.8%; Score 980; DB 4; Length 319;
Best Local Similarity 93.7%; Pred. No. 4.4e-90;
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPKQRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKQRTKNTNRRPMDVDFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
QY 61 RRQPIPARQPOGRHWAQPGYPMPLVYNGECGAGWLLSPRGSRRPHGPNDRRRSRNLG 120
DB 61 RRQPIPARQPOGRHWAQPGYPMPLVYNGECGAGWLLSPRGSRRPHGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 16
US-08-974-690C-217
Sequence 217, Application US/08974690C
Patent No. 661333
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
TITLE OF INVENTION: VIRUS
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: Patentin version 3.1
SEQ ID NO 217
LENGTH: 319
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-974-690C-217

Query Match 93.8%; Score 980; DB 4; Length 319;
Best Local Similarity 93.7%; Pred. No. 4.4e-90;
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPKQRTKNTNRRPPTDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
DB 1 MSTNPKQRTKNTNRRPMDVDFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
QY 61 RRQPIPARQPOGRHWAQPGYPMPLVYNGECGAGWLLSPRGSRRPHGPNDRRRSRNLG 120
DB 61 RRQPIPARQPOGRHWAQPGYPMPLVYNGECGAGWLLSPRGSRRPHGPNDRRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

Db 1 MSTNPKRQKTKNTNRPMVDVKEPGGGQIVGGVYLLPRRGPRILGVARTKTSERSQPRG 60
QY 61 RROPIDKARPOGRHMAOPGYPMPLVGNCGWAGMLSPRGSRRPHMGPNDRRRSRNLG 120
Db 61 RROPIDKARPOGRHMAOPGYPMPLVGNCGWAGMLSPRGSRRPHMGPNDRRRSRNLG 120
QY 121 KYIDTLTCGFADMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KYIDTLTCGFADMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 17
US-08-290-665A-195
Sequence 195, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUCH, J., MILLER, R.H. AND
PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 26
US-08-290-665A-195

Query Match 93.6%; Score 978; DB 2; Length 191;
Best Local Similarity 93.2%; Pred. No. 3.7e-90;
Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPKORTKTKNTNRPTDVKEPGGGQIVGGVYLLPRGPPLGVARTKTSERSQPRG 60
Db 1 MSTNPKRQKTKNTNRPMVDVKEPGGGQIVGGVYLLPRGPPLGVARTKTSERSQPRG 60
QY 61 RROPIDKARPOGRHMAOPGYPMPLVGNCGWAGMLSPRGSRRPHMGPNDRRRSRNLG 120

Db 61 RROPIDKARPSRGRSMAOPGYPMPLVGNCGWAGMLSPRGSRRPHMGPNDRRRSRNLG 120
QY 121 KYIDTLTCGFADMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KYIDTLTCGFADMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 18
PCT-US95-10398-195
Sequence 195, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUCH, J., MILLER, R.H. AND
PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 26
PCT-US95-10398-195

Query Match 93.6%; Score 978; DB 5; Length 191;
Best Local Similarity 93.2%; Pred. No. 3.7e-90;
Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPKORTKTKNTNRPTDVKEPGGGQIVGGVYLLPRGPPLGVARTKTSERSQPRG 60
Db 1 MSTNPKRQKTKNTNRPMVDVKEPGGGQIVGGVYLLPRGPPLGVARTKTSERSQPRG 60
QY 61 RROPIDKARPOGRHMAOPGYPMPLVGNCGWAGMLSPRGSRRPHMGPNDRRRSRNLG 120

Db 61 RQPIPKARSEGRSQAQPGYPMPLVGNBCCGMAWMLSPRGSRPSPMGPDPRRSRNLG 120
121 KYIDTLTCGFADLMGYIPVVGAPLGVAAALAHGVRALEDGINVATGNLPGCSFSIFLLA 180
QY 121 KYIDTLTCGFADLMGYIPVVGAPLGVAAALAHGVRALEDGINVATGNLPGCSFSIFLLA 180
Db 121 KYIDTLTCGFADLMGYIPVVGAPLGVAAALAHGVRALEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 19
US-08-290-665A-196
; Sequence 196, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 196:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z7
; US-08-290-665A-196

Query Match 93.5%; Score 977; DB 2; Length 191;
Best Local Similarity 93.2%; Pred. No. 4.7e-90;
Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPQRTKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPLGVRAATKTSERSQPRG 60
1 MSTNPKPQRTKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPLGVRAATKTSERSQPRG 60
Db 1 MSTNPKPQRTKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPLGVRAATKTSERSQPRG 60
QY 61 RROPITKARQPOGRHQAQPGYPMPLVGNBCCGMAWMLSPRGSRPSPMGPDPRRSRNLG 120
61 RROPITKARQPOGRHQAQPGYPMPLVGNBCCGMAWMLSPRGSRPSPMGPDPRRSRNLG 120
Db 61 RROPITKARQPOGRHQAQPGYPMPLVGNBCCGMAWMLSPRGSRPSPMGPDPRRSRNLG 120
QY 121 KYIDTLTCGFADLMGYIPVVGAPLGVAAALAHGVRALEDGINVATGNLPGCSFSIFLLA 180

Db 121 KYIDTLTCGFADLMGYIPVVGAPLGVAAALAHGVRALEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 20
PCT-US95-10398-196
; Sequence 196, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 196:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z7
; PCT-US95-10398-196

Query Match 93.5%; Score 977; DB 5; Length 191;
Best Local Similarity 93.2%; Pred. No. 4.7e-90;
Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPPQRTKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPLGVRAATKTSERSQPRG 60
1 MSTNPKPQRTKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPLGVRAATKTSERSQPRG 60
Db 1 MSTNPKPQRTKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPLGVRAATKTSERSQPRG 60
QY 61 RROPITKARQPOGRHQAQPGYPMPLVGNBCCGMAWMLSPRGSRPSPMGPDPRRSRNLG 120
61 RROPITKARQPOGRHQAQPGYPMPLVGNBCCGMAWMLSPRGSRPSPMGPDPRRSRNLG 120
Db 61 RROPITKARQPOGRHQAQPGYPMPLVGNBCCGMAWMLSPRGSRPSPMGPDPRRSRNLG 120

QY 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 21

US-08-836-075A-42
Sequence 42, Application US/08836075A
Patent No. 6180768

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/836,075A

FILING DATE: 21 Apr 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP95/04155

FILING DATE: 23 Oct 1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: INNS:004

FORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 319 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-836-075A-42

Query Match 93.5%; Score 977; DB 3; Length 319;

Best Local Similarity 93.7%; Pred. No. 8.8e-90;

Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPRKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
Db 1 MSTLPRKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60

QY 61 RROPPIKAROPQGRHMAOPGYPMPLYGNECGMAGWLLSPRGSPPHMGPNDRRRSRLG 120
Db 61 RROPPIKAROPQGRHMAOPGYPMPLYGNECGMAGWLLSPRGSPPHMGPNDRRRSRLG 120

QY 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

RESULT 22

US-08-635-886C-229
Sequence 229, Application US/08635886C
Patent No. 655114

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, Geert

APPLICANT: DELEYS, Robert

APPLICANT: MAERTENS, Geert

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

TITLE OF INVENTION: VIRUS

FILE REFERENCE: 2752-18

CURRENT APPLICATION NUMBER: US/08/635,886C

CURRENT FILING DATE: 1996-04-25

PRIOR APPLICATION NUMBER: PCT/EP94/03555

PRIOR FILING DATE: 1994-10-28

PRIOR APPLICATION NUMBER: EP 93402718.6

PRIOR FILING DATE: 1993-11-04

NUMBER OF SEQ ID NOS: 286

SOFTWARE: PatentIn version 3.1

SEQ ID NO 229

LENGTH: 319

TYPE: PR

ORGANISM: hepatitis C virus

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (128)..(129)

OTHER INFORMATION: Xaa is any amino acid

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (308)..(308)

OTHER INFORMATION: Xaa is any amino acid

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (310)..(310)

OTHER INFORMATION: Xaa is any amino acid

US-08-635-886C-229

Query Match 93.5%; Score 977; DB 4; Length 319;

Best Local Similarity 93.7%; Pred. No. 8.8e-90;

Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPRKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
Db 1 MSTLPRKQRTKNTNRPTDVKPPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60

QY 61 RROPPIKAROPQGRHMAOPGYPMPLYGNECGMAGWLLSPRGSPPHMGPNDRRRSRLG 120
Db 61 RROPPIKAROPQGRHMAOPGYPMPLYGNECGMAGWLLSPRGSPPHMGPNDRRRSRLG 120

QY 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDITLTCGFMADMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 229
LENGTH: 319
TYPE: PRT
ORGANISM: hepatitis C virus
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (128)..(129)
OTHER INFORMATION: xaa is any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (308)..(308)
OTHER INFORMATION: xaa is any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (310)..(310)
OTHER INFORMATION: xaa is any amino acid
US-08-974-690C-229

Query Match 93.5%; Score 977; DB 4; Length 319;
Best Local Similarity 93.7%; Pred. No. 8.8e-90;
Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RROPPIKAROPQGRHWAQGYEWPPLYGNECCGAGWLLSPRGSRPWPGNDPRRRSRNLG 120
DB 61 RROPPIKAROPQGRHWAQGYEWPPLYGNECCGAGWLLSPRGSRPWPGNDPRRRSRNLG 120
QY 121 KVDTLTGCFADIMGYIPVVGAPLGVAALAHGVRALIDGIVYATGNLPGCSFSIFLLA 180
DB 121 KVDTLTGCFADIMGYIPVVGAPLGVAALAHGVRALIDGIVYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 24
US-08-635-886C-181
Sequence 181, Application US/08635886C
Patent No. 6555114
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2752-18
CURRENT APPLICATION NUMBER: US/08/635,886C
CURRENT FILING DATE: 1996-04-25
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 181
LENGTH: 450
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-635-886C-181

Query Match 93.5%; Score 977; DB 4; Length 450;

Best Local Similarity 92.1%; Pred. No. 1.3e-89;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RROPPIKAROPQGRHWAQGYEWPPLYGNECCGAGWLLSPRGSRPWPGNDPRRRSRNLG 120
DB 61 RROPPIKAROPQGRHWAQGYEWPPLYGNECCGAGWLLSPRGSRPWPGNDPRRRSRNLG 120
QY 121 KVDTLTGCFADIMGYIPVVGAPLGVAALAHGVRALIDGIVYATGNLPGCSFSIFLLA 180
DB 121 KVDTLTGCFADIMGYIPVVGAPLGVAALAHGVRALIDGIVYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 25

US-08-974-690C-181
Sequence 181, Application US/08974690C
Patent No. 661333
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 181
LENGTH: 450
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-974-690C-181

Query Match 93.5%; Score 977; DB 4; Length 450;
Best Local Similarity 92.1%; Pred. No. 1.3e-89;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RROPPIKAROPQGRHWAQGYEWPPLYGNECCGAGWLLSPRGSRPWPGNDPRRRSRNLG 120
DB 61 RROPPIKAROPQGRHWAQGYEWPPLYGNECCGAGWLLSPRGSRPWPGNDPRRRSRNLG 120
QY 121 KVDTLTGCFADIMGYIPVVGAPLGVAALAHGVRALIDGIVYATGNLPGCSFSIFLLA 180
DB 121 KVDTLTGCFADIMGYIPVVGAPLGVAALAHGVRALIDGIVYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 26
US-08-466-975A-23
Sequence 23, Application US/08466975A
Patent No. 5910404
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
APPLICANT: POLLET, DIRK

APPLICANT: MAERTENS, GEERT
APPLICANT: VAN HEUVERSWUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,975A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/391,671
FILING DATE:
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
-466-975A-23

Query Match 93.5%; Score 977; DB 2; Length 2894;
Best Local Similarity 92.1%; Pred. No. 1.3e-88;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPPDVKFPGGQIVGVYLLPRGPRLGVRATKTSERSQPRG 60
DB 1 MSTLPKQRTKNTNRPPDVKFPGGQIVGVYLLPRGPRLGVRATKTSERSQPRG 60

QY 61 RROPPIKAROPGRHMAOPGYPMPLVGNBCCGNAWMLSPRSGRPHMGPNDRPRRSNLTG 120
DB 61 RROPPIKAROPGRHMAOPGYPMPLVGNBCCGNAWMLSPRSGRPHMGPNDRPRRSNLTG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 27
US-08-391-671A-23

Sequence 23, Application US/08391671A
Patent No. 5922532
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
APPLICANT: POLLET, DIRK
APPLICANT: MAERTENS, GEERT
APPLICANT: VAN HEUVERSWUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/391,671A
FILING DATE: 21-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-391-671A-23

Query Match 93.5%; Score 977; DB 2; Length 2894;
Best Local Similarity 92.1%; Pred. No. 1.3e-88;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKNTNRPPDVKFPGGQIVGVYLLPRGPRLGVRATKTSERSQPRG 60
DB 1 MSTLPKQRTKNTNRPPDVKFPGGQIVGVYLLPRGPRLGVRATKTSERSQPRG 60

QY 61 RROPPIKAROPGRHMAOPGYPMPLVGNBCCGNAWMLSPRSGRPHMGPNDRPRRSNLTG 120
DB 61 RROPPIKAROPGRHMAOPGYPMPLVGNBCCGNAWMLSPRSGRPHMGPNDRPRRSNLTG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 28
US-08-467-902A-23
Sequence 23, Application US/08467902A
Patent No. 6007982
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
APPLICANT: POLETT, DIRK
APPLICANT: MAERTENS, GERT
APPLICANT: VAN HEUVERSMUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
NUMBER OF SEQUENCES: 23
TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,902A
FILING DATE: US/08/467,902A
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/391,671
FILING DATE:
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA: WO PCT/EP91/02409
APPLICATION NUMBER: 13-DEC-1991
APPLICATION DATA: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-467-902A-23

Query Match 93.5%; Score 977; DB 3; Length 2894;
Best Local Similarity 92.1%; Pred. No. 1.3e-88;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
DB 1 MSTLPKQKTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
QY 61 RROPPIKAROPGRHMAQPGYPMPLVYNGECGMAAGWLLSPRGRSPHMGPNDRRRSRNLG 120
DB 61 RROPPIKAROPGRHMAQPGYPMPLVYNGECGMAAGWLLSPRGRSPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTGGFADLMGTYPLVVGAPLGGVAAALAHGVRAIEDGICINATGMLGCGCSSTFLA 180
DB 121 KVIDTLTGGFADLMGTYPLVVGAPLGGVAAALAHGVRAIEDGICINATGMLGCGCSSTFLA 180

QY 181 LUSCLTTPASA 191
DB 181 LUSCLTTPASA 191

RESULT 29
US-09-275-265-23
Sequence 23, Application US/09275265
Patent No. 6287761
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
APPLICANT: POLETT, DIRK
APPLICANT: MAERTENS, GERT
APPLICANT: VAN HEUVERSMUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
NUMBER OF SEQUENCES: 23
TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/275,265
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/391,671
FILING DATE: 21-FEB-1995
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA: WO PCT/EP91/02409
APPLICATION NUMBER: 13-DEC-1991
APPLICATION DATA: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-09-275-265-23

Query Match 93.5%; Score 977; DB 3; Length 2894;
Best Local Similarity 92.1%; Pred. No. 1.3e-88;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
DB 1 MSTLPKQKTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
QY 61 RROPPIKAROPGRHMAQPGYPMPLVYNGECGMAAGWLLSPRGRSPHMGPNDRRRSRNLG 120
DB 61 RROPPIKAROPGRHMAQPGYPMPLVYNGECGMAAGWLLSPRGRSPHMGPNDRRRSRNLG 120

Db 61 RROPPIKVRREBGRRTWAQPGYPMPLYGNBCCGAGWLLSPRGSRPWMGPDPRRSRNLG 120
QY 121 KVIDTLTCGFADIMGYIPLVGAPLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADIMGYIPLVGAPLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 30

US-09-941-611-23
Sequence 23, Application US/09941611
Patent No. 6576417

GENERAL INFORMATION:

APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWIN, HUGO

TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF

ANTIBODIES TO HEPATITIS C VIRUS

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match

Best Local Similarity 92.1%; Score 977; DB 4; Length 2894;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQRTKTKNTNRPTDVKFPFGGQIVGVYLLPRGPRILGVATRTKTSERSOPRG 60
DB 1 MSTLPKQRTKTKNTNRPTDVKFPFGGQIVGVYLLPRGPRILGVATRTKTSERSOPRG 60

QY 61 RROPPIKVRREBGRRTWAQPGYPMPLYGNBCCGAGWLLSPRGSRPWMGPDPRRSRNLG 120
Db 61 RROPPIKVRREBGRRTWAQPGYPMPLYGNBCCGAGWLLSPRGSRPWMGPDPRRSRNLG 120
QY 121 KVIDTLTCGFADIMGYIPLVGAPLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADIMGYIPLVGAPLGGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

Search completed: May 12, 2004, 09:17:05
Job time : 25 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2004, 09:11:55 ; Search time 39 Seconds
(without alignments)
1545.230 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 1045
Sequence: 1 MSTLPPQKRTKNTNRPT.....CSFSIFLMLSLCTTPASA 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 90 summaries

Database :

SPTREMBL_25:*
1: sp archaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp invertebrate:*
6: sp mammal:*
7: sp mhc:*
8: sp organelle:*
9: sp phage:*
10: sp plant:*
11: sp rodent:*
12: sp virus:*
13: sp vertebrate:*
14: sp unclassified:*
15: sp_rv1rus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1045	100.0	191	12	Q68115 hepatitis c
2	1035	99.0	191	12	Q68411 hepatitis c
3	1025	98.1	414	12	P89956 hepatitis c
4	1024	98.0	414	12	P89958 hepatitis c
5	1024	98.0	414	12	P89958 hepatitis c
6	1024	98.0	3018	12	O92531 hepatitis c
7	1023	97.9	3018	12	O39527 hepatitis c
8	1020	97.6	414	12	P89963 hepatitis c
9	1016	97.2	414	12	P89957 hepatitis c
10	1016	97.2	414	12	P89960 hepatitis c
11	1014	97.0	414	12	P89959 hepatitis c
12	1011	96.7	414	12	Q681270 hepatitis c
13	1009	96.6	319	12	Q68703 hepatitis c
14	1009	96.6	414	12	P89962 hepatitis c
15	1007	96.4	191	12	O93057 hepatitis c
16	1007	96.4	414	12	Q81269 hepatitis c

17	1003	96.0	414	12	Q81268 hepatitis c
18	1003	96.0	414	12	Q81314 hepatitis c
19	1001	95.8	191	12	O70635 hepatitis c
20	1000	95.7	414	12	Q81271 hepatitis c
21	999	95.6	414	12	P89955 hepatitis c
22	997	95.4	414	12	P89954 hepatitis c
23	997	95.4	415	12	P89961 hepatitis c
24	997	95.4	415	12	P89964 hepatitis c
25	997	95.4	415	12	P89965 hepatitis c
26	997	95.4	3015	12	O92532 hepatitis c
27	996	95.3	191	12	Q68153 hepatitis c
28	995	95.2	415	12	Q81315 hepatitis c
29	995	95.2	415	12	Q81272 hepatitis c
30	990	94.7	415	12	P87751 hepatitis c
31	990	94.7	3013	12	O92530 hepatitis c
32	988	94.5	414	12	Q81267 hepatitis c
33	988	94.5	3019	12	O92529 hepatitis c
34	987	94.4	191	12	P87841 hepatitis c
35	987	94.4	415	12	Q81274 hepatitis c
36	985	94.3	416	12	O39647 hepatitis c
37	984	94.2	319	12	Q68698 hepatitis c
38	984	94.2	415	12	Q81273 hepatitis c
39	982	94.0	191	12	Q68152 hepatitis c
40	982	94.0	191	12	Q68157 hepatitis c
41	982	94.0	319	12	Q68705 hepatitis c
42	981	93.9	191	12	Q68576 hepatitis c
43	981	93.9	191	12	Q68116 hepatitis c
44	981	93.9	326	12	Q81257 hepatitis c
45	981	93.9	3010	12	Q90368 hepatitis c
46	980	93.8	191	12	Q68109 hepatitis c
47	980	93.8	419	12	Q68704 hepatitis c
48	979	93.7	411	12	Q81812 hepatitis c
49	978	93.6	191	12	Q68155 hepatitis c
50	977	93.5	191	12	Q68156 hepatitis c
51	977	93.5	513	12	Q81220 hepatitis c
52	977	93.5	3010	12	Q90364 hepatitis c
53	977	93.5	3010	12	O09796 hepatitis c
54	977	93.5	3011	12	Q03463 hepatitis c
55	976	93.4	191	12	Q68149 hepatitis c
56	976	93.4	191	12	Q68113 hepatitis c
57	976	93.4	191	12	Q68123 hepatitis c
58	976	93.4	191	12	Q68139 hepatitis c
59	976	93.4	191	12	Q68124 hepatitis c
60	976	93.4	191	12	Q817M4 hepatitis c
61	976	93.4	191	12	Q88B51 hepatitis c
62	976	93.4	415	12	O995G6 hepatitis c
63	976	93.4	415	12	O995G1 hepatitis c
64	976	93.4	415	12	O995G7 hepatitis c
65	976	93.4	415	12	O995G4 hepatitis c
66	976	93.4	415	12	O995G5 hepatitis c
67	976	93.4	746	12	Q817M2 hepatitis c
68	976	93.4	778	12	Q04184 hepatitis c
69	976	93.4	3008	12	O39929 hepatitis c
70	976	93.4	3011	12	O91FE5 hepatitis c
71	976	93.4	3011	12	O36579 hepatitis c
72	976	93.4	3011	12	O36610 hepatitis c
73	976	93.4	3011	12	O92LS8 hepatitis c
74	976	93.4	3011	12	O36609 hepatitis c
75	976	93.4	3011	12	O36608 hepatitis c
76	975	93.3	191	12	Q8V828 hepatitis c
77	975	93.3	191	12	Q8V7X3 hepatitis c
78	975	93.3	191	12	Q8V816 hepatitis c
79	975	93.3	191	12	Q8V817 hepatitis c
80	975	93.3	191	12	Q8V824 hepatitis c
81	975	93.3	191	12	Q8V819 hepatitis c
82	975	93.3	191	12	Q8V763 hepatitis c
83	975	93.3	191	12	Q8V822 hepatitis c
84	975	93.3	191	12	Q8V818 hepatitis c
85	975	93.3	191	12	Q8V821 hepatitis c
86	975	93.3	191	12	Q8V814 hepatitis c
87	975	93.3	191	12	Q8V786 hepatitis c
88	975	93.3	191	12	Q8V823 hepatitis c
89	975	93.3	191	12	Q817P0 hepatitis c

90 975 93.3 191 12 Q8V829

Q8V829 hepatitis c

ALIGNMENTS

RESULT 1

Q68115 PRELIMINARY; PRT; 191 AA.
ID 068115
AC 068115;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Genome polyprotein (Fragment).
GN C.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK2;
RX MEDLINE=94336721; PubMed=8056787;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the core gene of 14 hepatitis C virus
genotypes.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HK2;
RX MEDLINE=92279243; PubMed=1317578;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the 5' noncoding region of hepatitis C virus.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HK2;
RX MEDLINE=93376778; PubMed=8396266;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "At least 12 genotypes of hepatitis C virus predicted by sequence
analysis of the putative E1 gene of isolates collected worldwide.";
RT Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN C AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; U0198; AAA21037.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR SO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON TER 191
SQ SEQUENCE 191 AA; 20725 MW; 7A7994DF2F0909C6 CRC64;

Query Match 100.0%; Score 1045; DB 12; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.7e-87;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLVGATRTKTSERSQPRG 60
DB 1 MSTLPKPKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLVGATRTKTSERSQPRG 60
QY 61 RROPIPKAROPQGRHWAQPGYPWPLVGNCGGAGWLLSPRGRPHWGPNDPRRSRLNG 120
DB 61 RROPIPKAROPQGRHWAQPGYPWPLVGNCGGAGWLLSPRGRPHWGPNDPRRSRLNG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 2

Q68411 PRELIMINARY; PRT; 191 AA.
ID 068411
AC 068411;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 6.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=42182;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QC26;
RX MEDLINE=97052554; PubMed=8897188;
RA Bernier L., Williams B., Delage G., Murphy D.G.;
RT "Identification of numerous hepatitis C virus genotypes in Montreal,
Canada.";
RT J. Clin. Microbiol. 34:2815-2818(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN C AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; U33435; AAB40038.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR SO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON TER 191
SQ SEQUENCE 191 AA; 20789 MW; F292AB64B56DE30A CRC64;

Query Match 99.0%; Score 1035; DB 12; Length 191;
Best Local Similarity 99.0%; Pred. No. 1.4e-86;
Matches 189; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSTLPKPKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLVGATRTKTSERSQPRG 60
DB 1 MSTLPKPKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLVGATRTKTSERSQPRG 60
QY 61 RROPIPKAROPQGRHWAQPGYPWPLVGNCGGAGWLLSPRGRPHWGPNDPRRSRLNG 120
DB 61 RROPIPKAROPQGRHWAQPGYPWPLVGNCGGAGWLLSPRGRPHWGPNDPRRSRLNG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 3

P89956 PRELIMINARY; PRT; 414 AA.
ID P89956
AC P89956;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;

RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN506;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN506;
 RA MEDLINE=95062197; PubMed=7972001;
 RX Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 RT Hepatitis C virus variants from Vietnam are classifiable into the
 RT seventh, eighth, and ninth major genetic groups.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; D88469; BAA13618.1; -
 DR PIR; P00804; P00804.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01539; HCV core; 1.
 DR Pfam; PF01560; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON TER
 SQ SEQUENCE 414 AA; 44528 MW; 7A3BE1710311C017 CRC64;
 Query Match 98.1%; Score 1025; DB 12; Length 414;
 Best Local Similarity 97.9%; Pred. No. 2.7e-85;
 Matches 187; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSTLPPORRTKNTNRRPTDVKFPGGGQIVGGVYLLPRRGLGVATRTKTSERSQPRG 60
 DB 1 MSTLPPORRTKNTNRRPTDVKFPGGGQIVGGVYLLPRRGLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKAROPGCRHMAOGYPMPLVGNBCCGAGWLLSPRGSRPWMGPNDPRRSRLNG 120
 DB 61 RROPPIKAROPGCRHMAOGYPMPLVGNBCCGAGWLLSPRGSRPWMGPNDPRRSRLNG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 4
 Q81329 PRELIMINARY; PRT; 414 AA.
 ID 081329
 AC 081329;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_Taxid=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN405;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN405;
 RA MEDLINE=95062197; PubMed=7972001;
 RX Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 RT Hepatitis C virus variants from Vietnam are classifiable into the
 RT seventh, eighth, and ninth major genetic groups.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; D88469; BAA13617.1; -
 DR PIR; P00804; P00804.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01539; HCV core; 1.
 DR Pfam; PF01560; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON TER
 SQ SEQUENCE 414 AA; 44937 MW; 65355640863B3DCF CRC64;
 Query Match 98.0%; Score 1024; DB 12; Length 414;
 Best Local Similarity 98.4%; Pred. No. 3.3e-85;
 Matches 188; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSTLPPORRTKNTNRRPTDVKFPGGGQIVGGVYLLPRRGLGVATRTKTSERSQPRG 60
 DB 1 MSTLPPORRTKNTNRRPTDVKFPGGGQIVGGVYLLPRRGLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKAROPGCRHMAOGYPMPLVGNBCCGAGWLLSPRGSRPWMGPNDPRRSRLNG 120
 DB 61 RROPPIKAROPGCRHMAOGYPMPLVGNBCCGAGWLLSPRGSRPWMGPNDPRRSRLNG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 5
 P89958 PRELIMINARY; PRT; 414 AA.
 ID P89958
 AC P89958;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_Taxid=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN530;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

```

RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-VN530;
RX MEDLINE=95062197; PubMed=7972001;
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Vietnam are classifiable into the
RT sevencl, elchit, and ninth major genetic groups.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D88471; BAA13620.1; -.
CC GO; GO:0016021; C:integral to membrane; IEA.
CC GO; GO:0019028; C:viral capsid; IEA.
CC GO; GO:0019031; C:viral envelope; IEA.
CC GO; GO:0005198; F:structural molecule activity; IEA.
CC InterPro; IPR002522; HCV_core.
CC InterPro; IPR002521; HCV_capsid.
CC InterPro; IPR002531; HCV_env.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NSI; 1.
CC Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW polypeptide; Transmembrane.
FT NON_TER
SQ SEQUENCE 414 AA; 414 MW; F3F3CF154372FAFF CRC64;

Query Match 98.0%; Score 1024; DB 12; Length 414;
Best Local Similarity 97.4%; Pred. No. 3, 3e-85;
Matches 186; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKNTNRPTDYKPGGGQIVGGVYLLPRRGLGVRAIRKTSERSQPRG 60
DB 1 MSTLPKQKTKNTNRPTDYKPGGGQIVGGVYLLPRRGLGVRAIRKTSERSQPRG 60

QY 61 RROPPIKAROPGRHMAOPGYPMPLYGNECCGAGMLSPRSGRPWGPNDPRRSRNIG 120
DB 61 RROPPIKAROPGRHMAOPGYPMPLYGNECCGAGMLSPRSGRPWGPNDPRRSRNIG 120

QY 121 KYIDTLTCGFADMGYIPVVGAPLGGVAAALAHGVAIEDGIVATGNLPGCSFSIFLTA 180
DB 121 KYIDTLTCGFADMGYIPVVGAPLGGVAAALAHGVAIEDGIVATGNLPGCSFSIFLTA 180

DB 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 6
092531 PRELIMINARY; PRT; 3016 AA.
ID 092531;
AC 092531;
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepicivirus.
OX NCB1_Taxid=11103;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-VN405;
RA Okamoto H.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN-VN405;
MEDLINE=98378034; PubMed=9714232;

```

```

RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F., Miyakawa Y.,
RA Mayumi M.;
RT "The entire nucleotide sequences of three hepatitis C virus isolates
RT in genetic groups 7-9 and comparison with those in the other eight
RT genetic groups.";
RL J. Gen. Virol. 79:1847-1857(1998).
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D84264; BAA32666.1; -.
CC F1R; PQ0804; PQ0804.
CC HSP; P27958; 1A1V.
CC MEROPS; S29.001; -.
CC GO; GO:0016021; C:integral to membrane; IEA.
CC GO; GO:0019028; C:viral capsid; IEA.
CC GO; GO:0019031; C:viral envelope; IEA.
CC GO; GO:0005524; F:ATP binding; IEA.
CC GO; GO:0008026; F:ATP dependent helicase activity; IEA.
CC GO; GO:0016787; F:hydrolyase activity; IEA.
CC GO; GO:0003723; F:RNA binding; IEA.
CC GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
CC GO; GO:0008236; F:serine-type peptidase activity; IEA.
CC GO; GO:0005198; F:structural molecule activity; IEA.
CC GO; GO:0016740; F:translational activity; IEA.
CC GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
CC GO; GO:0006350; P:transcription; IEA.
CC GO; GO:0019079; P:viral genome replication; IEA.
CC GO; GO:0019087; P:viral transformation; IEA.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NSI.
CC InterPro; IPR002518; HCV_NSI.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS5a.
CC InterPro; IPR002868; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRp.
CC InterPro; IPR004109; Peptidase_C29.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR007094; RNA_pol_Psitv.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NSI; 1.
CC Pfam; PF01538; HCV_NSI; 1.
CC Pfam; PF02907; HCV_NSI; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4a; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF00998; Viral_RdRp; 1.
CC Pfam; PF00998; Viral_RdRp; 1.
CC SMART; SM00487; DEXDC; 1.
CC ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3016 AA; 328031 MW; 455CF96258BC3B CRC64;

Query Match 98.0%; Score 1024; DB 12; Length 3016;
Best Local Similarity 98.4%; Pred. No. 3, 1e-84;
Matches 186; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKNTNRPTDYKPGGGQIVGGVYLLPRRGLGVRAIRKTSERSQPRG 60
DB 1 MSTLPKQKTKNTNRPTDYKPGGGQIVGGVYLLPRRGLGVRAIRKTSERSQPRG 60

QY 61 RROPPIKAROPGRHMAOPGYPMPLYGNECCGAGMLSPRSGRPWGPNDPRRSRNIG 120
DB 61 RROPPIKAROPGRHMAOPGYPMPLYGNECCGAGMLSPRSGRPWGPNDPRRSRNIG 120

```

QY 121 KVIDTLTCGADMGITPVVGAFLGVAAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGADMGITPVVGAFLGVAAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 7
 ID 039927 PRELIMINARY; PRT; 3018 AA.
 AC 039927;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus type 6a.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=31655;
 CC [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=euHk2;
 RX MEDLINE=97320431; PubMed=9177282;
 RA Adams A., Chamberlain R.W., Taylor L.A., Davidson F., Lin C.K.,
 RA Stimmone P., Elliot R.M.;
 RT "Complete coding sequence of hepatitis C virus genotype 6a."
 RL Biochem. Biophys. Res. Commun. 234:393-396(1997).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; Y12083; CAA72801.1; -.
 DR HSP; P27958; 1A1V.
 DR MEROPS; S29.001; -.
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:000368; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P: transcription; IEA.
 DR GO; GO:0019079; P: viral genome replication; IEA.
 DR GO; GO:0019087; P: viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_Lypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002531; HCV env.
 DR InterPro: IPR002518; HCV NS1.
 DR InterPro: IPR002515; HCV NS2.
 DR InterPro: IPR000745; HCV NS4A.
 DR InterPro: IPR002868; HCV NS4B.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR004109; Peptidase C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid. 1.
 DR Pfam: PF01542; HCV core. 1.
 DR Pfam: PF01539; HCV env. 1.
 DR Pfam: PF01560; HCV NS1. 1.
 DR Pfam: PF01538; HCV NS2. 1.
 DR Pfam: PF02907; HCV NS3. 1.
 DR Pfam: PF01006; HCV NS4a. 1.
 DR Pfam: PF01001; HCV NS4b. 1.

DR Pfam: PF01506; HCV NS5a. 1.
 DR Pfam: PF00998; Viral RdRp. 1.
 DR ProDom; PD186062; HCV NS1. 1.
 DR SMART; SM00487; DEXdc. 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3018 AA; 329016 MW; 6B67FB3CF6A61AE3 CRC64;
 Query Match 97.9%; Score 1023; DB 12; Length 3018;
 Best Local Similarity 97.9%; Pred. No. 3.8e-84;
 Matches 187; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSTLPKPKRTKNTNRPRPDVAFPGGGOIVGGVYLLPRRGRPLGRATKTEERQPRG 60
 DB 1 MSTLPKPKRTKNTNRPRPDVAFPGGGOIVGGVYLLPRRGRPLGRATKTEERQPRG 60
 QY 61 RROPTEKARQPGRRHQAQGYPMPLVGNCGAGWMLSPSGRPHMGPNDRRRSRLTG 120
 DB 61 RROPTEKARQPGRRHQAQGYPMPLVGNCGAGWMLSPSGRPHMGPNDRRRSRLTG 120
 QY 121 KVIDTLTCGADMGITPVVGAFLGVAAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGADMGITPVVGAFLGVAAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 8
 ID P89963 PRELIMINARY; PRT; 414 AA.
 AC P89963;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11103;
 CC [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=VN571;
 RX Okamoto H.;
 RN Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=VN571;
 RX MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Vietnam are classifiable into the
 RT seventh, eighth, and ninth major genetic groups."
 RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; D88476; BAA13625.1; -.
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR Pfam: PF01543; HCV capsid. 1.
 DR Pfam: PF01542; HCV core. 1.
 DR Pfam: PF01539; HCV env. 1.
 DR Pfam: PF01560; HCV NS1. 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polyprotein; Transmembrane.
 FT NON TER 414 414
 SQ SEQUENCE 414 AA; 4480 MW; 397DDDC5C4BDA20D CRC64;
 Query Match 97.6%; Score 1020; DB 12; Length 414;
 Best Local Similarity 96.9%; Pred. No. 7.7e-85;
 Matches 185; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSTLPKPKTKNTKNTNRRTDVKFPGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60
 DB 1 MSTLPKPKTKNTKNTNRRTDVKFPGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60
 QY 61 RROPFKARQPOGRHNAQPGYPMPLVGNCGAGMLSPRSGRPHWGPNDRRRSRNLG 120
 DB 61 RROPFKARQPOGRHNAQPGYPMPLVGNCGAGMLSPRSGRPHWGPNDRRRSRNLG 120
 QY 121 KVITDITCGFADLMGYIPVIGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVITDITCGFADLMGYIPVIGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 9
 P89957 PRELIMINARY; PRT; 414 AA.
 AC P89957;
 DT 01-MAY-1997 (TEMBLrel. 03, Created)
 DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN507;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN507;
 MEDLINE=95062197; Pubmed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 "Hepatitis C virus variants from Vietnam are classifiable into the
 seventh, eighth, and ninth major genetic groups.";
 Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; D88470; BAA13619.1; -
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR InterPro: IPR002519; HCV_core.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON TER 414 414
 SQ SEQUENCE 414 AA; 4480 MW; 1754DDDS83F52414 CRC64;

Query Match 97.2%; Score 1016; DB 12; Length 414;
 Best Local Similarity 96.9%; Pred. No. 1.8e-84;
 Matches 185; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSTLPKPKTKNTKNTNRRTDVKFPGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60
 DB 1 MSTLPKPKTKNTKNTNRRTDVKFPGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60
 QY 61 RROPFKARQPOGRHNAQPGYPMPLVGNCGAGMLSPRSGRPHWGPNDRRRSRNLG 120
 DB 61 RROPFKARQPOGRHNAQPGYPMPLVGNCGAGMLSPRSGRPHWGPNDRRRSRNLG 120
 QY 121 KVITDITCGFADLMGYIPVIGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVITDITCGFADLMGYIPVIGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 10
 P89960 PRELIMINARY; PRT; 414 AA.
 AC P89960;
 DT 01-MAY-1997 (TEMBLrel. 03, Created)
 DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN538;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN538;
 MEDLINE=95062197; Pubmed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 "Hepatitis C virus variants from Vietnam are classifiable into the
 seventh, eighth, and ninth major genetic groups.";
 Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; D88473; BAA13622.1; -
 DR PIR; P00804; P00804.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON TER 414 414
 SQ SEQUENCE 414 AA; 44630 MW; C7DDC84D0198BBD7 CRC64;
 Query Match 97.2%; Score 1016; DB 12; Length 414;
 Best Local Similarity 96.9%; Pred. No. 1.8e-84;
 Matches 185; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy	1	MSTLPKPKRKTKNNRRPTDVKPFPGGGQIVGVYLLPRRGRLGVRATRTKTSSESGORG	60
Dd	1	MSTLPKPKRKTKNNRRPMDVKKFPGGQIVGVYLLPRRGRLGVRATRTKTSSESGORG	60
Qy	61	RROPPIKAROPQGRHMAOPGYPMPLVYGNECGGMALLSPRGSRPDHMGPNDRRSRNIG	120
Dd	61	RROPPIKAROSQGRHMAOPGYPMPLVYNGECGMALMSPROSRPHMGPNDRHSRNIG	120
Qy	121	KVIDTLTGCFADLMKGITIVVGAPPLAGVAALAHGRALEDGINATGNLPGCSFSIFLLA	180
Dd	121	KVIDTLTGCFADLMGXIEIVVGAPPLAGVAALAHGRAVEDGINATGNLPGCSFSIFLLA	180
Qy	181	LISCCTTPASA 191	
Dd	181	LISCCTTPASA 191	
 RESULT 11 PRELIMINARY; PRT; 414 AA.			
A	898959	PRELIMINARY;	PRT; 414 AA.
DT	01-MAY-1997	(TREMBLrel. 03, Created)	
DT	01-MAY-1997	(TREMBLrel. 03, Last sequence update)	
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)	
DE	Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).		
OS	Hepatitis C virus.		
OC	Virusess; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OX	Hepadnavirus.		
NCBI	_taxid=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VN531;		
RA	Okamoto H.,		
RL	Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VN531;		
RX	MEDLINE=95062197; PubMed=7972001;		
RY	EMBL=D848472; BAI13621.1; -		
GO	GO:0016021; C:integral to membrane; IEA.		
GO	GO:0019028; C:viral capsid; IEA.		
GO	GO:0019031; C:viral envelope; IEA.		
GO	GO:0005198; F:structural molecule activity; IEA.		
InterPro	IPRO02522; HCV_capsid.		
InterPro	IPRO02521; HCV_core.		
InterPro	IPRO02519; HCV_env.		
InterPro	IPRO02531; HCV_NS1.		
pfam	PF01543; HCV_capsid; 1.		
pfam	PF01542; HCV_core; 1.		
pfam	PF01539; HCV_env; 1.		
pfam	PF01560; HCV_NS1; 1.		
KEGG	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;		
KW	Polypeptide; Transmembrane.		
FT	NON TER		
SO	SEQUENCE		
AA:	414	MM:	656500C405A2C082 CRC64;
 Query Match 97.0%; Score 1014; DB 12; Length 414; Best Local Similarity 96.9%; Pred. No. 2.7e-84; Matches 185; Conservative 3; Mismatches 3; Indels 0; Gaps 0;			
Qy	1	MSTLPKPKRKTKNNRRPTDVKPFPGGGQIVGVYLLPRRGRLGVRATRTKTSSESGORG	60
Dd	1	MSTLPKPKRKTKNNRRPMDVKKFPGGQIVGVYLLPRRGRLGVRATRTKTSSESGORG	60
Qy	61	RROPPIKAROPQGRHMAOPGYPMPLVYGNECGGMALLSPRGSRPDHMGPNDRRSRNIG	120
Dd	61	RROPPIKAROSQGRHMAOPGYPMPLVYNGECGMALMSPROSRPHMGPNDRHSRNIG	120
Qy	121	KVIDTLTGCFADLMKGITIVVGAPPLAGVAALAHGRALEDGINATGNLPGCSFSIFLLA	180
Dd	121	KVIDTLTGCFADLMGXIEIVVGAPPLAGVAALAHGRAVEDGINATGNLPGCSFSIFLLA	180
Qy	181	LISCCTTPASA 191	
Dd	181	LISCCTTPASA 191	

QY	61	PROPIPKARQPOGRHMAOPGYWMPLYNMGCGMAGMLSPGSRPHMGPNDRRSRNIG	120
Db	61	RRPIPKARQPTGRHMAOPGYWMPLYNMGCGMAGMLSPGSRPHMGPNDRRSRNIG	120
QY	121	KVIDTLTTCGFDLMGXIYIPVGAAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA	180
Db	121	KVIDITTCGVDLMGXIIYIPVLGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA	180
QY	181	LLSLCITTPASA 191	
Db	181	LLSLCITTPASA 191	
RESULT 12			
Q81270	Q81270	PRELIMINARY; PRT; 414 AA.	
AC	Q81270;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)		
DE	Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
CC	Hepacivirus.		
OX	NCBI_Taxid=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=7h552;		
RX	MEDLINE=96005057; PubMed=7561773;		
RA	Tokita H., Okamoto H., Luengrojjanakul P., Varesangthip K.,		
RA	Chaiyavut T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;		
RT	"Hepatitis C virus variants from Thailand classifiable into five novel		
RT	genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major		
RT	genetic groups."		
RL	J. Gen. Virol. 76:2329-2335 (1995).		
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A		
CC	LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:		
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF		
CC	PROTEIN C AND MNA (BY SIMILARITY).		
DR	EMBL, D37845; BAA07091.1; -		
DR	GO; GO:0016021; C:Integral to membrane; IEA.		
DR	GO; GO:0019028; C:viral capsid; IEA.		
DR	GO; GO:0019031; C:viral envelope; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	InterPro; IPR002522; HCV_capsid.		
DR	InterPro; IPR002521; HCV_core.		
DR	InterPro; IPR002519; HCV_env.		
DR	InterPro; IPR002531; HCV_NS1.		
DR	Pfam; PF01543; HCV_capsid; 1.		
DR	Pfam; PF01542; HCV_core; 1.		
DR	Pfam; PF01539; HCV_env; 1.		
DR	Pfam; PF01560; HCV_NS1; 1.		
DR	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;		
KW	Polypeptide; Transmembrane.		
KW	Polymer		
FT	NON TER		
FT	414		
SEQUENCE	414 AA; 44994 MW; 8C1E123FC61D89FC CRC64;		
Query Match	96.7%; Score 1011; DB 12; Length 414;		
Best Local Similarity	97.4%; Freq. No. 5, 1e-84;		
Matches	186; Conservative 0; Mismatches 5; Indels 0; Gaps 0		
QY	1	MSTLIPKPKRKRTNRRNRPPTVTKFPFGGQIYGVVYLPRRGPRILGVARTKTSERSQPG	60
Db	1	MSTLIPKPKRKRTNRRNRPPTVTKFPFGGQIYGVVYLPRRGPRILGVARTKTSERSQPG	60
QY	61	RRPIPKARQPOGRHMAOPGYWMPLYNMGCGMAGMLSPGSRPHMGPNDRRSRNIG	120
Db	61	RRPIPKARQPTGRHMAOPGYWMPLYNMGCGMAGMLSPGSRPHMGPNDRRSRNIG	120
QY	121	KVIDTLTTCGFDLMGXIYIPVGAAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA	180
Db	121	KVIDITTCGVDLMGXIIYIPVLGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA	180

QY 181 LLSCLTTPASA 191
 |||||
 DB 181 LLSCLTTPASA 191

RESULT 13

068703 PRELIMINARY; PRT; 319 AA.

AC 068703; 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, last sequence update)
 DT 01-JUN-2003 (TRENBLREL. 24, last annotation update)
 DE Core protein (Fragment).
 OS Hepatitis C virus type 6a.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 NC Hepacivirus.
 RN NCBI_Taxid=31655;
 [1]
 SEQUENCE FROM N.A.

RA MEDLINE=96118171; PubMed=857885;
 RA Stuyver L., Waseur A., van Arnhem W., Lunel F., Laurent-Puig P.,
 RA Pawlotsky J.-M., Kleter B., Bassit L., Nkengasong J., van Doorn L.-J.,
 RA Maertens G.;
 RT "Hepatitis C virus genotyping by means of 5'-UTR/core line probe assays
 RT and molecular analysis of untypeable samples.";
 RL Virus Res. 38:137-157(1995).
 DR EMBL; U38339; AAC42193.1; -.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0019031; C:Viral envelope; IEA.
 DR GO; GO:0005198; F:Structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 FT CHAIN 1 >191 CORE PROTEIN.
 FT NON_TER 319 >319 E1 PROTEIN.
 FT SEQUENCE 319 AA; 34545 MW; DC1D0EA52ED648F CRC64;

Query Match 96.6%; Score 1009; DB 12; Length 319;
 Best Local Similarity 96.3%; Pred. No. 5.8e-84;
 Matches 184; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSTLPPQKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
 |||||
 DB 1 MSTLPPQKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
 QY 61 RROPPIKARQPOGRHMAOQGYPMPLVGNBCCGAGWLLSPGSRPHMGPNDRRRSRNLG 120
 |||||
 DB 61 RROPPIKARQPOGRHMAOQGYPMPLVGNBCCGAGWLLSPGSRPHMGPNDRRRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 |||||
 DB 181 LLSCLTTPASA 191

RESULT 14

068962 PRELIMINARY; PRT; 414 AA.

AC 068962; 01-MAY-1997 (TRENBLREL. 03, Created)
 DT 01-MAY-1997 (TRENBLREL. 03, last sequence update)
 DT 01-JUN-2003 (TRENBLREL. 24, last annotation update)
 DE Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN569;
 RA Okamoto H.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RX MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Vietnam are classifiable into the
 RT seventh, eighth, and ninth major genetic groups.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL; D88475; BAAL3624.1; -.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0019031; C:Viral envelope; IEA.
 DR GO; GO:0005198; F:Structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polypeptide; Transmembrane.
 FT CHAIN 414 >414
 FT NON_TER 414 >414
 FT SEQUENCE 414 AA; 44751 MW; 4BBA61438D11E3C6 CRC64;

Query Match 96.6%; Score 1009; DB 12; Length 414;
 Best Local Similarity 96.3%; Pred. No. 7.8e-84;
 Matches 184; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSTLPPQKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
 |||||
 DB 1 MSTLPPQKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60
 QY 61 RROPPIKARQPOGRHMAOQGYPMPLVGNBCCGAGWLLSPGSRPHMGPNDRRRSRNLG 120
 |||||
 DB 61 RROPPIKARQPOGRHMAOQGYPMPLVGNBCCGAGWLLSPGSRPHMGPNDRRRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 |||||
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 |||||
 DB 181 LLSCLTTPASA 191

RESULT 15

093057 PRELIMINARY; PRT; 191 AA.

AC 093057; 068747; 01-NOV-1998 (TRENBLREL. 08, Created)
 DT 01-NOV-1998 (TRENBLREL. 08, last sequence update)
 DT 01-JUN-2003 (TRENBLREL. 24, last annotation update)
 DE Genome polypeptide (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 NC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=D97/93;
 RA MEDLINE=97060095; PubMed=8904416;
 RA Doi H., Apichartiyakul C., Ohba K., Mizokami M., Hotta H.;
 RT "Hepatitis C virus (HCV) subtype prevalence in Chiang Mai, Thailand,
 RT and identification of novel subtypes of HCV major types 6.";
 RL J. Clin. Microbiol. 34:569-574(1996).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: D63946; BAA09976.1; -
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Polyprotein.
 KW NON_TER
 SQ SEQUENCE 191 AA; 20727 MW; 411C1C851ADA2DB5 CRC64;
 Query Match 96.4%; Score 1007; DB 12; Length 191;
 Best Local Similarity 95.3%; Pred. No. 5e-84;
 Matches 182; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MSTLPKPKRKTNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVARATKTSERSQPRG 60
 Db 1 MSTLPKPKRKTNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVARATKTSERSQPRG 60
 QY 61 RROPPIKARQPOGRHMAQPGYPMWPLYGNECGAGWMLSPRGSRPHWGPNDRRRSNTLG 120
 Db 61 RROPPIKARQPOGRHMAQPGYPMWPLYGNECGAGWMLSPRGSRPHWGPNDRRRSNTLG 120
 QY 121 KVITDLTCGPADLMGTYIPVVGAPLGGVAAALAHGVAIEGIVYATGNLPGCSFSIFLLA 180
 Db 121 KVITDLTCGPADLMGTYIPVVGAPLGGVAAALAHGVAIEGIVYATGNLPGCSFSIFLLA 180
 QY 121 RVIDITTCGADLMGTYIPVVGAPLGGVAAALAHGVAIEGIVYATGNLPGCSFSIFLLA 180
 Db 121 RVIDITTCGADLMGTYIPVVGAPLGGVAAALAHGVAIEGIVYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191
 RESULT 16
 Q81269 PRELIMINARY; PRT; 414 AA.
 AC 081269;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 OS Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TH271;
 RX MEDLINE=96005057; PubMed=7561773;
 RA Tokita H., Okamoto H., Luengrojjanakul P., Vareesangthip K.,
 RA Chaiwattana T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Thailand classifiable into five novel
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genetic groups.";
 RL J. Gen. Virol. 76:2329-2335(1995).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: D37844; BAA07090.1; -
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.

DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002521; HCV env.
 DR InterPro: IPR002523; HCV NS1.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 KW NON_TER
 SQ SEQUENCE 414 AA; 45072 MW; 6C9E1AE6462241D CRC64;
 Query Match 96.4%; Score 1007; DB 12; Length 414;
 Best Local Similarity 96.9%; Pred. No. 1.2e-83;
 Matches 185; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MSTLPKPKRKTNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVARATKTSERSQPRG 60
 Db 1 MSTLPKPKRKTNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVARATKTSERSQPRG 60
 QY 61 RROPPIKARQPOGRHMAQPGYPMWPLYGNECGAGWMLSPRGSRPHWGPNDRRRSNTLG 120
 Db 61 RROPPIKARQPOGRHMAQPGYPMWPLYGNECGAGWMLSPRGSRPHWGPNDRRRSNTLG 120
 QY 121 KVITDLTCGPADLMGTYIPVVGAPLGGVAAALAHGVAIEGIVYATGNLPGCSFSIFLLA 180
 Db 121 KVITDLTCGPADLMGTYIPVVGAPLGGVAAALAHGVAIEGIVYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191
 RESULT 17
 Q81268 PRELIMINARY; PRT; 414 AA.
 AC 081268;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 OS Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TH846;
 RX MEDLINE=96005057; PubMed=7561773;
 RA Tokita H., Okamoto H., Luengrojjanakul P., Vareesangthip K.,
 RA Chaiwattana T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Thailand classifiable into five novel
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genetic groups.";
 RL J. Gen. Virol. 76:2329-2335(1995).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: D37843; BAA07089.1; -
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002521; HCV env.
 DR InterPro: IPR002521; HCV NS1.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV NS1; 1.

KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON TER 414 414
 SQ SEQUENCE 414 AA; 44836 MW; 212740491A9DA0B1 CRC64;

Query Match 96.0%; Score 1003; DB 12; Length 414;
 Best Local Similarity 96.9%; Pred. No. 2.8e-83;
 Matches 185; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRAIRKTSERSQPRG 60
 DB 1 MSTLPKPKRTKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRAIRKTSERSQPRG 60
 QY 61 RROPPIKARPOGGRHMAOPGYPMPLVGNCGAGWMLSPRGRSPHMGNDPRRRSRNIG 120
 DB 61 RROPPIKARPOGGRHMAOPGYPMPLVGNCGAGWMLSPRGRSPHMGNDPRRRSRNIG 120
 QY 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 18

081314 PRELIMINARY; PRT; 414 AA.

AC 081314;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=75171;
 RX MEDLINE=96005057; PubMed=7561773;
 RA Tokita H., Okamoto H., Iuengrojanakul P., Vareseangthip K.,
 Chaituvatt T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Thailand classifiable into five novel
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genetic groups.";
 F. J. Gen. Virol. 76:2329-2335 (1995).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; D38078; BAA07271.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002531; HCV_env.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON TER 414 414
 SQ SEQUENCE 414 AA; 45158 MW; A6DBF58491BE0843 CRC64;

Query Match 96.0%; Score 1003; DB 12; Length 414;
 Best Local Similarity 96.3%; Pred. No. 2.8e-83;
 Matches 184; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRAIRKTSERSQPRG 60
 DB 1 MSTLPKPKRTKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRAIRKTSERSQPRG 60

QY 61 RROPPIKARPOGGRHMAOPGYPMPLVGNCGAGWMLSPRGRSPHMGNDPRRRSRNIG 120
 DB 61 RROPPIKARPOGGRHMAOPGYPMPLVGNCGAGWMLSPRGRSPHMGNDPRRRSRNIG 120

QY 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 19

070635 PRELIMINARY; PRT; 191 AA.

AC 070635; O68744;
 DT 01-AUG-1998 (T-EMBLrel. 07, Created)
 DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B4/92;
 RX MEDLINE=97060095; PubMed=8904416;
 RA Doi H., Apichartpiyakul C., Ohba K., Mizokami M., Hotta H.;
 RT "Hepatitis C virus (HCV) subtype prevalence in Chiang Mai, Thailand,
 RT and identification of novel subtypes of HCV major types 6.";
 R. J. Clin. Microbiol. 34:569-574 (1996).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; D63943; BAA09974.1; -;
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 KW Polyprotein.
 FT NON TER 191 191
 SQ SEQUENCE 191 AA; 20738 MW; FD1967C0F4371C50 CRC64;

Query Match 95.8%; Score 1001; DB 12; Length 191;
 Best Local Similarity 95.8%; Pred. No. 1.8e-83;
 Matches 183; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRAIRKTSERSQPRG 60
 DB 1 MSTLPKPKRTKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRAIRKTSERSQPRG 60

QY 61 RROPPIKARPOGGRHMAOPGYPMPLVGNCGAGWMLSPRGRSPHMGNDPRRRSRNIG 120
 DB 61 RROPPIKARPOGGRHMAOPGYPMPLVGNCGAGWMLSPRGRSPHMGNDPRRRSRNIG 120

QY 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 20
081271 PRELIMINARY; PRT; 414 AA.

ID 081271
AC 081271; 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JUN-2003 (TREMblrel. 24, Last annotation update)
OS Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
OC Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TH976;
RX MEDLINE=96005057; PubMed=7561773;
RA Tokita H., Okamoto H., Luengrojankul P., Vareseangthip K.,
Chaituvatt T., Iizuka H., Tsuda P., Miyakawa Y., Mayumi M.,
"Hepatitis C virus variants from Thailand classifiable into five novel
genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
genetic groups.";
J. Gen. Virol. 76:2329-2335 (1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL: D37846; BAA07092.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01543; HCV_NSI.
DR Pfam: PF01542; HCV_capsid; 1.
DR Pfam: PF01539; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414 414
SQ SEQUENCE 414 AA; 45148 MW; D63BE7CED5B71776 CRC64;

Query Match 95.7%; Score 1000; DB 12; Length 414;
Best Local Similarity 96.3%; Pred. No. 5.2e-83;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTLPPQKTKNTNRRPTDVKFPGGQIVGGVYLLPRGPRPGVATRKTSERSQPRG 60
DB 1 MSTLPPQKTKNTNRRPTDVKFPGGQIVGGVYLLPRGPRPGVATRKTSERSQPRG 60

QY 61 RROPITKAROPGRHNAOPGYPMPLVGNCGCGAGMLSPRGRPHMGPNDRRRSRNLG 120
DB 61 RROPITKARHQRGRSNAOPGYPMPLVGNCGCGAGMLSPRGRPHMGPNDRRRSRNLG 120

QY 121 KVIDLTCGFADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTCGFADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 21
P89955 PRELIMINARY; PRT; 414 AA.

ID P89955
AC P89955; 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DE 01-JUN-2003 (TREMblrel. 24, Last annotation update)
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

DE Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VN085;
RX MEDLINE=95062197; PubMed=7972001;
RA Tokita H., Okamoto H., Tsuda P., Song P., Nakata S., Chose T.,
Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.,
"Hepatitis C virus variants from Vietnam are classifiable into the
seventh, eighth, and ninth major genetic groups.";
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026 (1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL: D88466; BAA13615.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414 414
SQ SEQUENCE 414 AA; 44890 MW; 50343390BA2663A CRC64;

Query Match 95.6%; Score 999; DB 12; Length 414;
Best Local Similarity 95.3%; Pred. No. 6.4e-83;
Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTLPPQKTKNTNRRPTDVKFPGGQIVGGVYLLPRGPRPGVATRKTSERSQPRG 60
DB 1 MSTLPPQKTKNTNRRPTDVKFPGGQIVGGVYLLPRGPRPGVATRKTSERSQPRG 60

QY 61 RROPITKAROPGRHNAOPGYPMPLVGNCGCGAGMLSPRGRPHMGPNDRRRSRNLG 120
DB 61 RROPITKAROPGRHNAOPGYPMPLVGNCGCGAGMLSPRGRPHMGPNDRRRSRNLG 120

QY 121 KVIDLTCGFADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTCGFADIMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 22
P89954 PRELIMINARY; PRT; 414 AA.

ID P89954
AC P89954; 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DE 01-JUN-2003 (TREMblrel. 24, Last annotation update)
OS Core, env and part of E2/NS1 (Genome polypeptide) (Fragment).
OC Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

NCBI_TaxID=11103;
[1]
SEQUENCE FROM N.A.
STRAIN=VN004;
Okamoto H.;
Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
STRAIN=VN004;
MEDLINE=95062197; PubMed=7972001;
Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;
"Hepatitis C virus variants from Vietnam are classifiable into the
seventh, eighth, and ninth major genetic groups.";
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: D88465; BAA13614.1; -
GO: GO:0016021; C:integral to membrane; IEA.
GO: GO:0019028; C:viral capsid; IEA.
GO: GO:0019031; C:viral envelope; IEA.
GO: GO:0005198; F:structural molecule activity; IEA.
InterPro: IPR002522; HCV_capsid.
InterPro: IPR002521; HCV_core.
InterPro: IPR002519; HCV_env.
InterPro: IPR002531; HCV_NSI.
Pfam: PF01543; HCV_capsid; 1.
Pfam: PF01542; HCV_core; 1.
Pfam: PF01539; HCV_env; 1.
Pfam: PF01560; HCV_NSI; 1.
ProDom: PD186062; HCV_NSI; 1.
Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; Transmembrane.
NON_TER
SEQUENCE 414 AA; 44560 MW; 373D93269DEB3E2 CRC64;
SQ
Query Match 95.4%; Score 997; DB 12; Length 414;
Best Local Similarity 95.3%; Pred. No. 9.7e-83;
Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTLPRQKRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLAGVRAATKTSERSQPRG 60
DB 1 MSTLPRQKRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLAGVRAATKTSERSQPRG 60
QY 61 RQPIPKAQPQGRHQAQGYPMPLTGNCGAGWLLSPRSRPMWGPNDPRRSRNIG 120
DB 61 RQPIPKAQPQGRHQAQGYPMPLTGNCGAGWLLSPRSRPMWGPNDPRRSRNIG 120
QY 121 KVIDITTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDITTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191
RESULT 23
P89961 PRELIMINARY; PRT; 415 AA.
ID P89961
AC P89961;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.

STRAIN=VN540;
Okamoto H.;
Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
STRAIN=VN540;
MEDLINE=95062197; PubMed=7972001;
Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;
"Hepatitis C virus variants from Vietnam are classifiable into the
seventh, eighth, and ninth major genetic groups.";
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: D88474; BAA13623.1; -
GO: GO:0016021; C:integral to membrane; IEA.
GO: GO:0019028; C:viral capsid; IEA.
GO: GO:0019031; C:viral envelope; IEA.
GO: GO:0005198; F:structural molecule activity; IEA.
GO: GO:0003700; F:transcription factor activity; IEA.
GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
InterPro: IPR002522; HCV_capsid.
InterPro: IPR002521; HCV_core.
InterPro: IPR002519; HCV_env.
InterPro: IPR002531; HCV_NSI.
InterPro: IPR000847; HTH_LYER.
Pfam: PF01543; HCV_capsid; 1.
Pfam: PF01542; HCV_core; 1.
Pfam: PF01539; HCV_env; 1.
Pfam: PF01560; HCV_NSI; 1.
ProSITE: PS00044; HTH_LYER_FAMILY; 1.
Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; Transmembrane.
NON_TER
SEQUENCE 415 AA; 44979 MW; B7C72222E6C46A32 CRC64;
SQ
Query Match 95.4%; Score 997; DB 12; Length 415;
Best Local Similarity 95.3%; Pred. No. 9.7e-83;
Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTLPRQKRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLAGVRAATKTSERSQPRG 60
DB 1 MSTLPRQKRTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLAGVRAATKTSERSQPRG 60
QY 61 RQPIPKAQPQGRHQAQGYPMPLTGNCGAGWLLSPRSRPMWGPNDPRRSRNIG 120
DB 61 RQPIPKAQPQGRHQAQGYPMPLTGNCGAGWLLSPRSRPMWGPNDPRRSRNIG 120
QY 121 KVIDITTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDITTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191
RESULT 24
P89964 PRELIMINARY; PRT; 415 AA.
ID P89964
AC P89964;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=VN787;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN787;
 RA MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tada F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.,
 RA "Hepatitis C virus variants from Vietnam are classifiable into the
 RT seventh, eighth, and ninth major genetic groups."
 RT Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; D88477; BAA13626.1; -
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro: IPR002522; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
 DR Polyprotein; Transmembrane.
 DR NON TER 415
 SO SEQUENCE 415 AA; 44771 MW; 45A5AE3F6E2431A5 CRC64;
 Query Match 95.4%; Score 997; DB 12; Length 415;
 Best Local Similarity 95.3%; Pred. No. 9, 7e-83;
 Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MSTLPPQKTKRNTNRRPVDYKFRGGQIVGGVILLPRGPRLGVRATRKTSERQPRG 60
 DB 1 MSTLPPQKTKRNTNRRPVDYKFRGGQIVGGVILLPRGPRLGVRATRKTSERQPRG 60
 QY 61 RROPIKAPQGRHMAQPGYPMPLVGNCGAGWLLSPRGSRPHMGNDPRRSRNIG 120
 DB 61 RROPIKAPQGRHMAQPGYPMPLVGNCGAGWLLSPRGSRPHMGNDPRRSRNIG 120
 QY 121 KYIDLTTCGFADLMGYIPVGAFLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYIDLTTCGFADLMGYIPVGAFLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 25
 ID P89965 PRELIMINARY; PRT; 415 AA.
 AC P89965;
 DT 01-MAY-1997 (TReMBLrel. 03, Created)
 DT 01-MAY-1997 (TReMBLrel. 03, last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)
 DE Core, env and part of E2/NS1 (Genome polyprotein) (fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN843;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=VN843;
 RX MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tada F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.,
 RA "Hepatitis C virus variants from Vietnam are classifiable into the
 RT seventh, eighth, and ninth major genetic groups."
 RT Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; D88478; BAA13627.1; -
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR000847; HTH_LYER.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR PROSITE; PS00044; HTH_LYER_FAMILY; 1.
 DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
 DR Polyprotein; Transmembrane.
 DR NON TER 415
 SO SEQUENCE 415 AA; 44784 MW; CB4DFA8BDBF0969 CRC64;
 Query Match 95.4%; Score 997; DB 12; Length 415;
 Best Local Similarity 95.3%; Pred. No. 9, 7e-83;
 Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MSTLPPQKTKRNTNRRPVDYKFRGGQIVGGVILLPRGPRLGVRATRKTSERQPRG 60
 DB 1 MSTLPPQKTKRNTNRRPVDYKFRGGQIVGGVILLPRGPRLGVRATRKTSERQPRG 60
 QY 61 RROPIKAPQGRHMAQPGYPMPLVGNCGAGWLLSPRGSRPHMGNDPRRSRNIG 120
 DB 61 RROPIKAPQGRHMAQPGYPMPLVGNCGAGWLLSPRGSRPHMGNDPRRSRNIG 120
 QY 121 KYIDLTTCGFADLMGYIPVGAFLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYIDLTTCGFADLMGYIPVGAFLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191
 RESULT 26
 ID O92532 PRELIMINARY; PRT; 3015 AA.
 AC O92532;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN004;
 RA Okamoto H.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=VN004;
 RX MEDLINE=98378034; PubMed=9714232;
 RA Tokita H., Okamoto H., Itzuka H., Kishimoto J., Tsuda F., Miyakawa Y.,
 RA Mayumi M.;
 RT "The entire nucleotide sequences of three hepatitis C virus isolates
 RT in genetic groups 7-9 and comparison with those in the other eight
 RT genetic groups.";
 RL J. Gen. Virol. 79:1847-1857 (1998).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; D84265; AAA32667.1; -.
 DR HSSP; P27958; 1HE1.
 DR MEROPS; S29.001; -.
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:0003868; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P: translation; IEA.
 DR GO; GO:0019079; P: viral genome replication; IEA.
 DR GO; GO:0019087; P: viral transformation; IEA.
 DR InterPro; IPR009003; Cys_Ser_cryptin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR003006; 1g_MHC.
 DR InterPro; IPR004109; Peptidase C29.
 DR InterPro; IPR007095; RNA pol DS_PS.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 DR PROSITE; PS00290; IG_MHC; 1.
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3015 AA; 327974 MW; 69108DD32B5DA012 CRC64;

Query Match 95.4%; Score 997; DB 12; Length 3015;
 Best Local Similarity 95.3%; Pred. No. 9e-82; Indels 0; Gaps 0;
 Matches 182; Conservative 3; Mismatches 6;

QY 1 MSTLPKQKTKKNTNRPTDVKFPGGQIVGGVYLLPRRGRLGVATRTKTSERSQPRG 60
 DB 1 MSTLPKQKTKKNTNRPTDVKFPGGQIVGGVYLLPRRGRLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKARQPGGRHAAQPGYPMPLVGNCGMAAGMLSPRSGRPHMGNDPPRRSRNLG 120

DB 61 RROPPIKARQPGGRHAAQPGYPMPLVGNCGMAAGMLSPRSGRPHMGNDPPRRSRNLG 120
 QY 121 KYIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRATEDGINATGNLPGCSFSIFLLA 180
 DB 121 KYIDTLTCGLADLMGYIPVVGAPLGGVAAALAHGVRATEDGINATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 27
 Q68153
 ID 068153 PRELIMINARY; PRT; 191 AA.
 AC 068153;
 DT 01-NOV-1996 (TREMELREL. 01, Created)
 DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
 DT 01-OCT-2003 (TREMELREL. 25, Last annotation update)
 DE Genome polyprotein (Fragment).
 GN C.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis virus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=24;
 RX MEDLINE=94336721; PubMed=8058787;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "Sequence analysis of the core gene of 14 hepatitis C virus
 RT genotypes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=24;
 RX MEDLINE=92279243; PubMed=1317578;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "Sequence analysis of the 5' noncoding region of hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946 (1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=24;
 RX MEDLINE=93376778; PubMed=8396266;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "At least 12 genotypes of hepatitis C virus predicted by sequence
 RT analysis of the putative E1 gene of isolates collected worldwide.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238 (1993).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; U0236; AAA21075.1; -.
 DR PIR; P00804; P00804.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 KM Polyprotein.
 FT NON_TER 191
 SQ SEQUENCE 191 AA; 20726 MW; 738DD912C6DEDC9 CRC64;

Query Match 95.8%; Score 996; DB 12; Length 191;
 Best Local Similarity 94.8%; Pred. No. 5e-83; Indels 0; Gaps 0;
 Matches 181; Conservative 4; Mismatches 6;

QY 1 MSTLPKQKTKKNTNRPTDVKFPGGQIVGGVYLLPRRGRLGVATRTKTSERSQPRG 60
 DB 1 MSTLPKQKTKKNTNRPTDVKFPGGQIVGGVYLLPRRGRLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKARQPGGRHAAQPGYPMPLVGNCGMAAGMLSPRSGRPHMGNDPPRRSRNLG 120

DT 01-MAY-1997 (TrEMBLrel. 03, last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
DE Core, env, and part of E2/NS1 (Genome polypeptide) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VN235;
RA Okamoto H.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VN235;
RX MEDLINE=95062197; PubMed=7972001;
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Vietnam are classifiable into the
RT seventh, eighth, and ninth major genetic groups.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
CC -1- SUBUNIT. THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN M AND MRNA (BY SIMILARITY).
CC EMBL; D88467; BAA13616.1; -
DR GO:0016021; C:integral to membrane; IEA.
DR GO:0019028; C:viral capsid; IEA.
DR GO:0019031; C:viral envelope; IEA.
DR GO:0005198; F:structural molecule activity; IEA.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR pfam: PF01543; HCV_capsid; 1.
DR pfam: PF01542; HCV_core; 1.
DR pfam: PF01539; HCV_env; 1.
DR pfam: PF01560; HCV_NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polypeptide; Transmembrane.
FT NON_TER
SQ SEQUENCE 415 AA; 45301 MW; 39B2040A86517D1B CRC64;

Query Match 94.7%; Score 990; DB 12; Length 415;
Best local Similarity 94.2%; Pred. No. 4.2e-82;
Matches 180; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Q 1 MSTLPKQKTKNTNRPTDVKPPGQGIIVGVYLLPRRPRLLGVATRTKTSERSOPRG 60
D 1 MSTLPKQKTKNTNRPTDVKPPGQGIIVGVYLLPRRPRLLGVATRTKTSERSOPRG 60
Q 61 RROPIKAPQPGRHMAOPGYPMWPLGNBEGCGMAGMLSPRGSRPHMGPNDRPRRSRLG 120
D 61 RROPIKAPQPGRHMAOPGYPMWPLGNBEGCGMAGMLSPRGSRPHMGPNDRPRRSRLG 120
Q 121 KVIDTLTCGADMGYIPVVGAPLGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
D 121 KVIDTLTCGADMGYIPVVGAPLGVAAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
Q 181 LLSCLTTPASA 191
D 181 LLSCLTTPASA 191

Search completed: May 12, 2004, 09:15:59
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2004, 09:12:31 ; Search time 21 Seconds

(without alignments)
874.885 Million cell updates/sec

Title: US-09-084-691B-206

Sequence: 1045
1 MSTLPRKQKTKNTNRPRP.....CSFSIFLALLSCLTTPASA 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 90 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	977	93.5	3011	1 S40770	genome polypprotein
2	976	93.4	3011	1 GNMVCH	genome polypprotein
3	971	92.9	441	2 S12707	genome polypprotein
4	969	92.7	782	2 S18031	genome polypprotein
5	969	92.7	787	2 PN0677	hypothetical prote
6	968	92.6	3011	1 GNMVCH	genome polypprotein
7	965	92.3	3010	1 GNMVCH	genome polypprotein
8	965	92.3	3010	1 GNMVCH	genome polypprotein
9	962	92.1	550	2 JH0711	genome polypprotein
10	961	92.0	640	2 JQ1584	genome polypprotein
11	959	91.8	189	2 S32740	genome polypprotein
12	958	91.7	782	2 S19876	genome polypprotein
13	957	91.6	533	2 PC1284	genome polypprotein
14	957	91.6	782	2 S19875	genome polypprotein
15	957	91.6	876	2 PC2219	polypeptide - hepa
16	956	91.5	3010	1 S18030	genome polypprotein
17	956	91.5	3010	1 GNMVCH	genome polypprotein
18	952	91.1	520	2 JQ1925	polypprotein - hepa
19	951	91.0	411	2 PC2060	genome polypprotein
20	948	90.7	3010	1 A45573	genome polypprotein
21	947	90.6	369	2 S21471	genome polypprotein
22	947	90.6	523	2 JQ1926	polypprotein - hepa
23	947	90.6	782	2 S18032	genome polypprotein
24	946	90.5	492	2 S41288	genome polypprotein
25	944	90.3	411	2 PC2061	genome polypprotein
26	941	90.0	874	2 JQ0883	genome polypprotein
27	940	90.0	513	2 A44150	structural protein
28	939	89.9	874	2 JQ0881	genome polypprotein
29	939	89.9	3033	1 JQ1303	genome polypprotein

30	936	89.6	3033	1 GNMVCH	genome polypprotein
31	919	87.8	178	2 PS0388	genome polypprotein
32	918	87.8	332	2 JN0265	genome polypprotein
33	893	85.5	3014	1 JCS620	genome polypprotein
34	647	61.9	124	2 S41360	genome polypprotein
35	635	60.8	123	2 S41361	genome polypprotein
36	631.5	60.4	266	2 PQ0393	genome polypprotein
37	626	59.9	118	2 S41346	genome polypprotein
38	617	59.0	114	2 S41362	genome polypprotein
39	612	58.6	117	2 S41363	genome polypprotein
40	611	58.5	115	2 S41345	genome polypprotein
41	610	58.4	115	2 S41364	genome polypprotein
42	609	58.3	114	2 S41365	genome polypprotein
43	605	57.9	114	2 S41366	genome polypprotein
44	605	57.9	115	2 S41344	genome polypprotein
45	604	57.8	114	2 S41358	genome polypprotein
46	601	57.5	115	2 S41343	genome polypprotein
47	600	57.4	114	2 S41359	genome polypprotein
48	597	57.1	115	2 S41342	genome polypprotein
49	597	57.1	115	2 S41351	genome polypprotein
50	596	57.0	112	2 S41341	genome polypprotein
51	596	57.0	114	2 S41370	genome polypprotein
52	595	56.9	115	2 S41347	genome polypprotein
53	594	56.8	125	2 S41352	genome polypprotein
54	593	56.7	115	2 S41350	genome polypprotein
55	592	56.7	114	2 S41368	genome polypprotein
56	591	56.6	115	2 S41349	genome polypprotein
57	586	56.1	114	2 S41369	genome polypprotein
58	578	55.3	112	2 S41371	genome polypprotein
59	561	53.7	109	2 S41367	genome polypprotein
60	551	52.7	108	2 S41348	genome polypprotein
61	550	52.6	115	2 S41354	genome polypprotein
62	549	52.5	108	2 S41390	genome polypprotein
63	547	52.3	108	2 S41357	genome polypprotein
64	540	51.7	108	2 S41353	genome polypprotein
65	539	51.6	108	2 S41356	genome polypprotein
66	537	51.4	108	2 S41355	genome polypprotein
67	419	40.1	82	2 PQ0804	core protein - hep
68	383	36.7	88	2 S21336	genome polypprotein
69	382	36.6	75	2 PQ0805	core protein - hep
70	226	21.6	180	2 PC1304	genome polypprotein
71	226	21.6	180	2 PC1303	genome polypprotein
72	225	21.5	180	2 PC1305	genome polypprotein
73	127	12.2	415	2 PC4407	envelope protein -
74	97.5	9.3	420	2 A45166	protein-lysine 6-O
75	94	9.0	348	2 A49594	enhancer factor pr
76	94	9.0	458	2 S12444	hypothetical prote
77	93.5	8.9	256	2 JU0268	major prion protei
78	93.5	8.9	256	2 S37149	prion protein - go
79	93.5	8.9	256	2 A54281	major prion protei
80	93.5	8.9	254	2 A54330	major prion protei
81	91.5	8.8	264	2 S37137	prion protein - gr
82	91	8.7	252	2 JC6175	hypothetical prote
83	90.5	8.7	628	2 S01955	probable secreted
84	89.5	8.5	114	2 D72600	hypothetical prote
85	89	8.5	252	2 S53634	major prion protei
86	88.5	8.5	252	2 G83022	probable two-compo
87	88.5	8.5	257	2 JQ1900	major prion protei
88	88	8.3	262	2 T14610	hypothetical prote
89	87	8.3	262	2 T14610	hypothetical prote
90	87	8.3	539	2 T28770	hypothetical prote

ALIGNMENTS

RESULT 1

S40770 genome polypprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (NS3, NS4, NS5A, NS5B, NS5C, NS5D, NS5E, NS5F, NS5G, NS5H, NS5I, NS5J, NS5K, NS5L, NS5M, NS5N, NS5O, NS5P, NS5Q, NS5R, NS5S, NS5T, NS5U, NS5V, NS5W, NS5X, NS5Y, NS5Z, NS5AA, NS5AB, NS5AC, NS5AD, NS5AE, NS5AF, NS5AG, NS5AH, NS5AI, NS5AJ, NS5AK, NS5AL, NS5AM, NS5AN, NS5AO, NS5AP, NS5AQ, NS5AR, NS5AS, NS5AT, NS5AU, NS5AV, NS5AW, NS5AX, NS5AY, NS5AZ, NS5BA, NS5BB, NS5BC, NS5BD, NS5BE, NS5BF, NS5BG, NS5BH, NS5BI, NS5BJ, NS5BK, NS5BL, NS5BM, NS5BN, NS5BO, NS5BP, NS5BQ, NS5BR, NS5BS, NS5BT, NS5BU, NS5BV, NS5BW, NS5BX, NS5BY, NS5BZ, NS5CA, NS5CB, NS5CC, NS5CD, NS5CE, NS5CF, NS5CG, NS5CH, NS5CI, NS5CJ, NS5CK, NS5CL, NS5CM, NS5CN, NS5CO, NS5CP, NS5CQ, NS5CR, NS5CS, NS5CT, NS5CU, NS5CV, NS5CW, NS5CX, NS5CY, NS5CZ, NS5DA, NS5DB, NS5DC, NS5DD, NS5DE, NS5DF, NS5DG, NS5DH, NS5DI, NS5DJ, NS5DK, NS5DL, NS5DM, NS5DN, NS5DO, NS5DP, NS5DQ, NS5DR, NS5DS, NS5DT, NS5DU, NS5DV, NS5DW, NS5DX, NS5DY, NS5DZ, NS5EA, NS5EB, NS5EC, NS5ED, NS5EE, NS5EF, NS5EG, NS5EH, NS5EI, NS5EJ, NS5EK, NS5EL, NS5EM, NS5EN, NS5EO, NS5EP, NS5EQ, NS5ER, NS5ES, NS5ET, NS5EU, NS5EV, NS5EW, NS5EX, NS5EY, NS5EZ, NS5FA, NS5FB, NS5FC, NS5FD, NS5FE, NS5FF, NS5FG, NS5FH, NS5FI, NS5FJ, NS5FK, NS5FL, NS5FM, NS5FN, NS5FO, NS5FP, NS5FQ, NS5FR, NS5FS, NS5FT, NS5FU, NS5FV, NS5FW, NS5FX, NS5FY, NS5FZ, NS5GA, NS5GB, NS5GC, NS5GD, NS5GE, NS5GF, NS5GG, NS5GH, NS5GI, NS5GJ, NS5GK, NS5GL, NS5GM, NS5GN, NS5GO, NS5GP, NS5GQ, NS5GR, NS5GS, NS5GT, NS5GU, NS5GV, NS5GW, NS5GX, NS5GY, NS5GZ, NS5HA, NS5HB, NS5HC, NS5HD, NS5HE, NS5HF, NS5HG, NS5HH, NS5HI, NS5HJ, NS5HK, NS5HL, NS5HM, NS5HN, NS5HO, NS5HP, NS5HQ, NS5HR, NS5HS, NS5HT, NS5HU, NS5HV, NS5HW, NS5HX, NS5HY, NS5HZ, NS5IA, NS5IB, NS5IC, NS5ID, NS5IE, NS5IF, NS5IG, NS5IH, NS5II, NS5IJ, NS5IK, NS5IL, NS5IM, NS5IN, NS5IO, NS5IP, NS5IQ, NS5IR, NS5IS, NS5IT, NS5IU, NS5IV, NS5IW, NS5IX, NS5IY, NS5IZ, NS5JA, NS5JB, NS5JC, NS5JD, NS5JE, NS5JF, NS5JG, NS5JH, NS5JI, NS5JJ, NS5JK, NS5JL, NS5JM, NS5JN, NS5JO, NS5JP, NS5JQ, NS5JR, NS5JS, NS5JT, NS5JU, NS5JV, NS5JW, NS5JX, NS5JY, NS5JZ, NS5KA, NS5KB, NS5KC, NS5KD, NS5KE, NS5KF, NS5KG, NS5KH, NS5KI, NS5KJ, NS5KK, NS5KL, NS5KM, NS5KN, NS5KO, NS5KP, NS5KQ, NS5KR, NS5KS, NS5KT, NS5KU, NS5KV, NS5KW, NS5KX, NS5KY, NS5KZ, NS5LA, NS5LB, NS5LC, NS5LD, NS5LE, NS5LF, NS5LG, NS5LH, NS5LI, NS5LJ, NS5LK, NS5LL, NS5LM, NS5LN, NS5LO, NS5LP, NS5LQ, NS5LR, NS5LS, NS5LT, NS5LU, NS5LV, NS5LW, NS5LX, NS5LY, NS5LZ, NS5MA, NS5MB, NS5MC, NS5MD, NS5ME, NS5MF, NS5MG, NS5MH, NS5MI, NS5MJ, NS5MK, NS5ML, NS5MM, NS5MN, NS5MO, NS5MP, NS5MQ, NS5MR, NS5MS, NS5MT, NS5MU, NS5MV, NS5MW, NS5MX, NS5MY, NS5MZ, NS5NA, NS5NB, NS5NC, NS5ND, NS5NE, NS5NF, NS5NG, NS5NH, NS5NI, NS5NJ, NS5NK, NS5NL, NS5NM, NS5NO, NS5NP, NS5NQ, NS5NR, NS5NS, NS5NT, NS5NU, NS5NV, NS5NW, NS5NX, NS5NY, NS5NZ, NS5OA, NS5OB, NS5OC, NS5OD, NS5OE, NS5OF, NS5OG, NS5OH, NS5OI, NS5OJ, NS5OK, NS5OL, NS5OM, NS5ON, NS5OO, NS5OP, NS5OQ, NS5OR, NS5OS, NS5OT, NS5OU, NS5OV, NS5OW, NS5OX, NS5OY, NS5OZ, NS5PA, NS5PB, NS5PC, NS5PD, NS5PE, NS5PF, NS5PG, NS5PH, NS5PI, NS5PJ, NS5PK, NS5PL, NS5PM, NS5PN, NS5PO, NS5PP, NS5PQ, NS5PR, NS5PS, NS5PT, NS5PU, NS5PV, NS5PW, NS5PX, NS5PY, NS5PZ, NS5QA, NS5QB, NS5QC, NS5QD, NS5QE, NS5QF, NS5QG, NS5QH, NS5QI, NS5QJ, NS5QK, NS5QL, NS5QM, NS5QN, NS5QO, NS5QP, NS5QQ, NS5QR, NS5QS, NS5QT, NS5QU, NS5QV, NS5QW, NS5QX, NS5QY, NS5QZ, NS5RA, NS5RB, NS5RC, NS5RD, NS5RE, NS5RF, NS5RG, NS5RH, NS5RI, NS5RJ, NS5RK, NS5RL, NS5RM, NS5RN, NS5RO, NS5RP, NS5RQ, NS5RR, NS5RS, NS5RT, NS5RU, NS5RV, NS5RW, NS5RX, NS5RY, NS5RZ, NS5SA, NS5SB, NS5SC, NS5SD, NS5SE, NS5SF, NS5SG, NS5SH, NS5SI, NS5SJ, NS5SK, NS5SL, NS5SM, NS5SN, NS5SO, NS5SP, NS5SQ, NS5SR, NS5SS, NS5ST, NS5SU, NS5SV, NS5SW, NS5SX, NS5SY, NS5SZ, NS5TA, NS5TB, NS5TC, NS5TD, NS5TE, NS5TF, NS5TG, NS5TH, NS5TI, NS5TJ, NS5TK, NS5TL, NS5TM, NS5TN, NS5TO, NS5TP, NS5TQ, NS5TR, NS5TS, NS5TT, NS5TU, NS5TV, NS5TW, NS5TX, NS5TY, NS5TZ, NS5UA, NS5UB, NS5UC, NS5UD, NS5UE, NS5UF, NS5UG, NS5UH, NS5UI, NS5UJ, NS5UK, NS5UL, NS5UM, NS5UN, NS5UO, NS5UP, NS5UQ, NS5UR, NS5US, NS5UT, NS5UU, NS5UV, NS5UW, NS5UX, NS5UY, NS5UZ, NS5VA, NS5VB, NS5VC, NS5VD, NS5VE, NS5VF, NS5VG, NS5VH, NS5VI, NS5VJ, NS5VK, NS5VL, NS5VM, NS5VN, NS5VO, NS5VP, NS5VQ, NS5VR, NS5VS, NS5VT, NS5VU, NS5VV, NS5VW, NS5VX, NS5VY, NS5VZ, NS5WA, NS5WB, NS5WC, NS5WD, NS5WE, NS5WF, NS5WG, NS5WH, NS5WI, NS5WJ, NS5WK, NS5WL, NS5WM, NS5WN, NS5WO, NS5WP, NS5WQ, NS5WR, NS5WS, NS5WT, NS5WU, NS5WV, NS5WX, NS5WY, NS5WZ, NS5XA, NS5XB, NS5XC, NS5XD, NS5XE, NS5XF, NS5XG, NS5XH, NS5XI, NS5XJ, NS5XK, NS5XL, NS5XM, NS5XN, NS5XO, NS5XP, NS5XQ, NS5XR, NS5XS, NS5XT, NS5XU, NS5XV, NS5XW, NS5XX, NS5XY, NS5XZ, NS5YA, NS5YB, NS5YC, NS5YD, NS5YE, NS5YF, NS5YG, NS5YH, NS5YI, NS5YJ, NS5YK, NS5YL, NS5YM, NS5YN, NS5YO, NS5YP, NS5YQ, NS5YR, NS5YS, NS5YT, NS5YU, NS5YV, NS5YW, NS5YX, NS5YY, NS5YZ, NS5ZA, NS5ZB, NS5ZC, NS5ZD, NS5ZE, NS5ZF, NS5ZG, NS5ZH, NS5ZI, NS5ZJ, NS5ZK, NS5ZL, NS5ZM, NS5ZN, NS5ZO, NS5ZP, NS5ZQ, NS5ZR, NS5ZS, NS5ZT, NS5ZU, NS5ZV, NS5ZW, NS5ZX, NS5ZY, NS5ZZ

C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_19-May-2000 #text_change 19-Jan-2001

C/Accession: S40770; PC1285
 R/Okamoto, H.
 Submitted to the EMBL Data Library, March 1992
 A/Reference number: S40770
 A/Accession: S40770
 A/Molecule type: genomic RNA
 A/Residues: 1-3011 <OKA>
 C/Cross-references: EMBL:DJ0749; NID:g221586; PIDN:BA01582.1; PID:g221587
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
 A/Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A/Reference number: PC1284; MUID:9101316; PMID:2170712
 A/Accession: PC1285
 A/Molecule type: genomic RNA
 A/Residues: 1-513 <OK2>
 A/Cross-references: GB:000831; NID:g221511; PIDN:BA00705.1; PID:g221512
 A/Experimental source: isolate HC-J1
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F/115/Product: capsid protein C #status predicted <CPC>
 F/192-191/Product: envelope protein M #status predicted <EMP>
 F/390-729/Product: major envelope protein E #status predicted <NEB>
 F/730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
 F/1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)
 F/1312-1317/Region: nucleotide-binding motif B
 F/1316-1319/Region: DEXH motif
 F/1616-1862/Product: nonstructural protein NS4 #status predicted <N4A>
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 93.5%; Score 977; DB 1; Length 3011;
 Best Local Similarity 92.1%; Pred. No. 1,7e-74;
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGVYLLPRRGRLGYRATRTKTSERQPRG 60
 DB 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGVYLLPRRGRLGYRATRTKTSERQPRG 60
 QY 61 RROPPIKARPOGRHNAOPGYPMPLVNGECGAGWLLSPGSRPHMGNDPDRSRNLG 120
 DB 61 RROPPIKARPOGRHNAOPGYPMPLVNGECGAGWLLSPGSRPHMGNDPDRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 2

GENVCH

genome polyprotein - hepatitis C virus (strain H)
 N/Contents: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C/Species: hepatitis C virus
 A/Note: host Homo sapiens (man)
 C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C/Accession: A36814; A41546
 R/Inchoupe, G.; Zebdeed, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to Genbank, July 1992
 A/Description: Genomic structure of the human prototype strain H of hepatitis C virus: C
 A/Reference number: A36814
 A/Accession: A36814
 A/Molecule type: genomic RNA
 A/Residues: 1-3011 <INC>
 A/Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
 R/Inchoupe, G.; Zebdeed, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A/Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar

A/Reference number: A41546; MUID:92052256; PMID:1658800
 A/Contents: annotation
 A/Note: neither amino acid nor nucleotide sequence is given
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F/115/Product: capsid protein C #status predicted <CPC>
 F/192-191/Product: envelope protein M #status predicted <EMP>
 F/390-729/Product: major envelope protein E #status predicted <NEB>
 F/730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
 F/1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)
 F/1312-1317/Region: nucleotide-binding motif B
 F/1316-1319/Region: DEXH motif
 F/1616-1862/Product: nonstructural protein NS4 #status predicted <N4A>
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F/196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,236

Query Match 93.4%; Score 976; DB 1; Length 3011;
 Best Local Similarity 92.7%; Pred. No. 2.1e-74;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGVYLLPRRGRLGYRATRTKTSERQPRG 60
 DB 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGVYLLPRRGRLGYRATRTKTSERQPRG 60
 QY 61 RROPPIKARPOGRHNAOPGYPMPLVNGECGAGWLLSPGSRPHMGNDPDRSRNLG 120
 DB 61 RROPPIKARPOGRHNAOPGYPMPLVNGECGAGWLLSPGSRPHMGNDPDRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 3

S12707

genome polyprotein - hepatitis C virus (fragment)
 N/Contents: core protein; envelope protein
 C/Species: hepatitis C virus
 C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
 C/Accession: S12707
 R/Takenuchi, K.; Kubo, Y.; Boonmat, S.; Watanabe, Y.; Katayama, T.; Choo, Q.L.; Kuo, G.; f
 Nucleic Acids Res. 18, 4626, 1990
 A/Title: Nucleotide sequence of core and envelope genes of the hepatitis C virus genome
 A/Reference number: S12707; MUID:90356432; PMID:2117749
 A/Accession: S12707
 A/Molecule type: genomic RNA
 A/Residues: 1-441 <TRK>
 A/Cross-references: EMBL:DO0574; NID:g221656; PIDN:BA00452.1; PID:g221657
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: polyprotein

Query Match 92.9%; Score 971; DB 2; Length 441;
 Best Local Similarity 92.7%; Pred. No. 9.1e-75;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGVYLLPRRGRLGYRATRTKTSERQPRG 60
 DB 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGVYLLPRRGRLGYRATRTKTSERQPRG 60
 QY 61 RROPPIKARPOGRHNAOPGYPMPLVNGECGAGWLLSPGSRPHMGNDPDRSRNLG 120
 DB 61 RROPPIKARPOGRHNAOPGYPMPLVNGECGAGWLLSPGSRPHMGNDPDRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVAIAEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 4

genome polyprotein - hepatitis C virus (isolate JK2) (fragment)
N/contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C/species: hepatitis C virus
A/Variety: isolate JK2
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C/Accession: S18031
R/Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL Data Library, September 1991
A/Description: Sequence analysis of putative structural regions of Hepatitis C Virus 18031
A/Reference number: S18029
A/Accession: S18031
A/Molecule type: genomic RNA
A/Residues: 1-782 <HON>
A/Cross-references: EMBL:X61593
C/Experimental source: isolate JK2
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match Best Local Similarity 92.7%; Score 969; DB 2; Length 782;

Matches 177; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRG 60
Db 1 MSTNPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRG 60
QY 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGAGWLLSPRGRSPHMGPNDRRRSRNLG 120
Db 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGAGWLLSPRGRSPHMGPNDRRRSRNLG 120
QY 121 KVIDTLTCGPADIMGYIPVVGAPLGGVAAALAHGVAIDGIVATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGPADIMGYIPVVGAPLGGVAAALAHGVAIDGIVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 5

genome polyprotein - hepatitis C virus (fragment)
N/contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C/species: hepatitis C virus
A/Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C/Accession: P00677
R/Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.
Biochem. Biophys. Res. Commun. 196, 780-788, 1993
A/Title: Genetic typing of hepatitis C viruses from Korean patients: Implications of gen
A/Reference number: P00677; PMID:94059104; PMID:8240354
A/Accession: P00677
A/Molecule type: mRNA
A/Residues: 1-787 <CHO>
A/Cross-references: GB:120498; NID:G1381031; PIDN:AA02608.1; PID:G1381032
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: glycoprotein; nonstructural protein
F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate

Query Match Best Local Similarity 92.7%; Score 969; DB 2; Length 787;
Matches 177; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRG 60
Db 1 MSTNPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRG 60

QY 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGAGWLLSPRGRSPHMGPNDRRRSRNLG 120
Db 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGAGWLLSPRGRSPHMGPNDRRRSRNLG 120

QY 121 KVIDTLTCGPADIMGYIPVVGAPLGGVAAALAHGVAIDGIVATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGPADIMGYIPVVGAPLGGVAAALAHGVAIDGIVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 6

genome polyprotein - hepatitis C virus (strain HCV-1)
N/contains: capsid protein C; envelope protein M; hepatitis virus (BC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/species: hepatitis C virus
C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C/Accession: A39166; P00404
R/Cho, O.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Col
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A/Title: Genetic organization and diversity of the hepatitis C virus.
A/Reference number: A39166; PMID:91172826; PMID:1848704
A/Accession: A39166

A/Molecule type: genomic RNA

A/Residues: 1-3011 <CHO>

A/Cross-references: GB:M62321; NID:G329873; PIDN:AAA5676.1; PID:G329874

A/Gen. S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.

J. Gen. Virol. 73, 1131-1141, 1992

A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to ex

A/Reference number: P00393; PMID:92268871; PMID:1316939

A/Accession: P00403

A/Molecule type: genomic RNA

A/Residues: 1577-1633 <GHA>

A/Cross-references: DDBJ:DI0128

A/Experimental source: Isolates E-b16

A/Status: preliminary

A/Molecule type: genomic RNA

A/Residues: 1577-1633 <CH2>

A/Experimental source: Isolates E-b17

C/Superfamily: hepatitis C virus genome polyprotein; glycoprotein; hydrolyase; nonstructural

C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolyase; nonstructural

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EMP>

F:192-389/Product: major envelope protein B #status predicted <MEB>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis virus #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXA motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22,

Query Match Best Local Similarity 92.6%; Score 968; DB 1; Length 3011;

Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRG 60
Db 1 MSTNPKPKRTKNTNRPTDVKFPGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRG 60
QY 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGAGWLLSPRGRSPHMGPNDRRRSRNLG 120
Db 61 RROPPIKARQPOGRHMAQPGYPMPLVGNCGGAGWLLSPRGRSPHMGPNDRRRSRNLG 120

F:191-380/Product: envelope protein E1 #status predicted <EPI>
 F:181-514/Product: envelope protein E2 or nonstructural protein NS1 (fragment) #status p
 F:115-550/Product: nonstructural protein NS5 (fragment) #status predicted <NS5>
 F:196-233,250,305,416,422,429,447/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match 92.1%; Score 962; DB 2; Length 550;
 Best Local Similarity 91.6%; Pred. No. 6,5e-74;
 Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 DB 1 MSTNPKPQKRTKNTNRPPQDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKARPOGRHMAQPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120
 DB 61 RROPPIKARPEGRRTAOPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120
 QY 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 10

JQ1584
 genome polypeptide - hepatitis C virus (strain U.K.) (fragment)
 N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural pro
 C:Species: hepatitis C virus
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
 C:Accession: JQ1584
 R:Kumar, U.; Thomas, H.; Monjardino, J.
 J. Gen. Virol. 73, 1521-1525, 1992
 A:Title: Cloning and sequencing of the structural region and expression of putative core
 A:Reference number: JQ1584; PMID:92300349; PMID:1318944
 A:Accession: JQ1584

A:Molecule type: genomic RNA
 A:Residues: 1-640 <RUM>
 A:Cross-references: GB:X84079; NID:G643119; PIDN:CAA5888.1; PID:G643120
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polype
 F:1-191/Product: core protein C #status predicted <CPC>
 F:192-389/Product: envelope protein E1 #status predicted <EPI>
 F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <
 F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova

Query Match 92.0%; Score 961; DB 2; Length 640;
 Best Local Similarity 91.1%; Pred. No. 9.1e-74;
 Matches 174; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 DB 1 MSTNPKPQKRTKNTNRPPQDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKARPOGRHMAQPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120
 DB 61 RROPPIKARPEGRRTAOPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120
 QY 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 11

S32740
 polypeptide - hepatitis C virus (isolate Russian) (fragment)
 N:Contains: capsid protein C; envelope protein M

C:Species: hepatitis C virus
 C:Date: 19-Mar-1997 #sequence_revision 05-Dec-1998 #text_change 17-Nov-2000
 C:Accession: S32740
 R:Vassilev, V.B.; Viazov, S.O.; Kotova, E.Y.; Nosikov, V.V.
 submitted to the EMBL Data Library, April 1993
 A:Description: Evidence of new HCV variant of European isolate in Russia.
 A:Reference number: S32740
 A:Accession: S32740

A:Molecule type: genomic RNA
 A:Residues: 1-189 <VAS>
 A:Cross-references: EMBL:X71407
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; envelope protein; polypeptide
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-189/Product: envelope protein M #status predicted <EPI>

Query Match 91.8%; Score 959; DB 2; Length 189;
 Best Local Similarity 92.1%; Pred. No. 4.3e-74;
 Matches 174; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 DB 1 MSTNPKPQKRTKNTNRPPQDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKARPOGRHMAQPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120
 DB 61 RROPPIKARPEGRRTAOPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120
 QY 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPA 189
 DB 181 LLSCLTTPA 189

RESULT 12

S19876
 genome polypeptide - hepatitis C virus (isolate JKS) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JKS
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19876
 R:Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A:Reference number: S18029
 A:Accession: S19876

A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61595; NID:G59486; PIDN:CAA43792.1; PID:G59487
 A:Experimental source: isolate JKS
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F:1-191/Product: core protein #status predicted <MAT1>
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 91.7%; Score 958; DB 2; Length 782;
 Best Local Similarity 91.6%; Pred. No. 2e-73;
 Matches 175; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 DB 1 MSTNPKPQKRTKNTNRPPQDVKFPGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60
 QY 61 RROPPIKARPOGRHMAQPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120
 DB 61 RROPPIKARPEGRRTAOPGYPMPLVGNCGCGAGWLLSPGSRPMPGPNDRPRRSRNIG 120

QY	121	KVIDITLTGCGFADLNGKXIPVVGAPLGGSVAALAHGVRAIEDGINVTGNLPGCSFSFI.LLA	180
		: : : :	
Dd	121	KVIDITLTGCGFADPMKXIPLVVGAFPLGGARALAHGVRVLDEDSVNATGNLPCCSFSFI.FLLA	180
		: : : :	
QY	181	LLSCLTTPASA 191	
Dd	181	LISCLTTPASA 191	

RESULT 13

PC1284

genome: polyprotein - hepatitis C virus (isolate HC-U4) (fragment)

C|Species: hepatitis C virus

C|Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000

C|Accession: PC1284

R|Okamoto, H.; Okada, S.; Sugiyama, Y.; Yoksumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda

ujun. J. Exp. Med. 60, 167-177, 1990

A|Title: The 5'-terminal sequence of the hepatitis C virus genome.

A|Reference number: PC1284; MUID:9103116; PMID:2170712

A|Accession: PC1284

A|Molecule type: genomic RNA

A|Residues: 1-513 <OK>

A|Cross-references: GB:DD0832; NID:g221513; PIDN:BA00706.1; PID:g221514

A|Superfamily: hepatitis C virus genome polyprotein

RESULT 14
S: 5
C: polyprotein - hepatitis C virus (isolate JK3) (fragment)
N: Nucleotides: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C: Species: hepatitis C virus
A: Variety: isolate JK3
C: Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C: Accession: SI9875
R: Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL Data Library, September 1991
A: Description: Sequence analysis of putative structural regions of Hepatitis C Virus isolate
A: Reference number: SI8029
A: Accession: SI9875
A: Molecule type: genomic RNA
A: Residues: 1-782 <HON>
A: Cross-references: EMBL:X61592; NID:G59482; PIDN:CAAA4789.1; PID:G59483
A: Experimental source: isolate JK3
C: Superfamily: hepatitis C virus genome polyprotein
C: Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F: 1-191/P: product: core protein #status predicted <MAT1>
F: 152-363/P: product: envelope protein 1 #status predicted <MAT2>
F: 354-733/P: product: NS1/E2 protein #status predicted <MAT3>
F: 734-782/P: product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Qy	1	MSLTPEKDKTKRNNRPRPTDVKPFGGGQIVGGVYLLPRRGPRLGVAATKTSRSRQ	60
Db	1	MSLNPRPQQTKNNTTRRPQDVKFPFGGQIVGGVYLLPRRGPRLGVAATKTSRSRQ	60
Qy	61	RRQPIPKAQPOGRHWAQGYPMPLYLENGCGGAGMLSPRSGRPHNGPNDPRRSNLTG	120
Db	61	RRPPIPKAQPEGRHWAQGYPMPLYLENGEGIGMTGMLSPRSGRPSMGPTDPRRSNLTG	120
Qy	121	KVIDTLTCGPADLMGTYIVVGAPPLCGVAALAHGVRALIEDGINATGNLPCCSFIPLA	180
Db	121	KVIDTLTTCGFADLMGTYIVVGAPPLCGVARALAHGVRVLEDEVNATGNLPCCSFIPLA	180
Qy	181	LISCLTTPASA	191
Db	181	LISCLTTPASA	191

RESULT 15
 PC2219
 polypeptide - hepatitis C virus (type 5a) (fragments)
 N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A protein
 C:Species: hepatitis C virus
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #ext_change 17-Nov-2000
 C:Accession: PC2219
 R:Stuyver, L.; Arnhem, W.V.; Wyeur, A.; Maertens, G.
 Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994
 A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the he
 A:Reference number: PC2219; MUID:94338342; PMID:7520237
 A:Accession: PC2219
 A:Molecule type: mRNA
 A:Residues: 1-876 <STU>
 A:Cross-references: GB:L29577; GB:L29578; GB:L29579
 A:Experimental source: serum
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: glycoprotein
 F:1-191/Product: core #status predicted <COE>
 F:68-78/Region: variable
 F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>
 F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>
 F:248-338/Region: E2
 F:339-411/Region: NS1 (amino end)
 F:412-783/Product: NS3 #status predicted <NSR>
 F:784-837/Product: NS4A #status predicted <NSB>
 F:838-876/Product: NS4B #status predicted <NSB>
 F:281, 287, 294, 312, 340/Binding site: carbohydrate (Asn) (covalent) #status predicted.

Query Match	Score 957, DB 2;	Length 782;
Best Local Similarity 91.6%;	Pred. No. 2,4e-73;	
Matches 175, Conservative 5;	Mismatches 11;	Indels 0; Gaps 0;

RESULT 16
S18030
genome polyprotein - hepatitis C virus (isolate Jk1)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructun
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus
 A:Virus: isolate JKI
 C:Date: 19-May-2000 #sequence revision 19-May-2000 #text_change 23-Mar-2001
 C:Accession: S18030; S33570; A48332; S18029
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient
 A:Reference number: S18028
 A:Accession: S18030
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <HOW>
 A:Cross-references: EMBL:X61596; NID:G59478; PID:CAA43793.1; PID:G59479
 A:Experimental source: isolate JKI from an individual
 R:Honda, M.; Kaneko, S.; Uonuma, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
 A:Reference number: A48332; MUID:93119270; PMID:8380332
 A:Accession: S33570
 A:Molecule type: genomic RNA
 A:Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>
 A:Cross-references: EMBL:X61591
 A:Note: This sequence is inconsistent with the nucleotide translation
 as TTP, and TTC for residue 771 as Ser
 A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:P.121748)
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <ME>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Product: hepatitis C virus predicted motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 91.5%; Score 956; DB 1; Length 3010;
 Best Local Similarity 91.6%; Pred. No. 1e-72;
 Matches 174; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLPRGPRLVGRATRTKTSERSOPRG 60
 N:Contains: C protein; E1 protein; E2/NS1 protein
 C:Species: hepatitis C virus
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000
 C:Accession: JQ1925
 R:Abde, K.; Inchausti, G.; Fujisawa, K.
 J. Gen. Virol. 73, 2725-2729, 1992
 A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated from a
 A:Reference number: JQ1925; MUID:93019030; PMID:1383400
 A:Accession: JQ1925
 A:Molecule type: RNA
 A:Residues: 1-520 <AB>
 A:Cross-references: DDBJ:U0687; NID:G221544; PID:BAA01529.1; PID:G221545
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: polypeptide; transmembrane protein
 F:1-191/Product: C protein #status predicted <CPR>
 F:192-383/Product: E1 protein #status predicted <E1>
 F:384-520/Product: E2/NS1 protein #status predicted <E2>

QY 1 MSTLPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLPRGPRLVGRATRTKTSERSOPRG 60
 DB 1 MSTNPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLTRGPRLVGRATRTKTSERSOPRG 60
 QY 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 DB 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPAS 190
 DB 181 LLSCLTTPAS 190
 QY 181 LLSCLTTPAS 190
 DB 181 LLSCLTTPAS 190

RESULT 17
 GNNVTV
 genome polypeptide - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (BC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40244
 R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
 Virology 188, 102-113, 1992

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449
 A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 A:Cross-references: GB:M84754
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <ME>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,250,305,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 91.5%; Score 956; DB 1; Length 3010;
 Best Local Similarity 91.6%; Pred. No. 1e-72;
 Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLPRGPRLVGRATRTKTSERSOPRG 60
 DB 1 MSTNPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLPRGPRLVGRATRTKTSERSOPRG 60
 QY 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 DB 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPAS 191
 DB 181 LLSCLTTPAS 191
 QY 181 LLSCLTTPAS 191
 DB 181 LLSCLTTPAS 191

RESULT 18
 JQ1925
 polypeptide - hepatitis C virus (isolate HCV-KP)
 N:Contains: C protein; E1 protein; E2/NS1 protein
 C:Species: hepatitis C virus
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000
 C:Accession: JQ1925
 R:Abde, K.; Inchausti, G.; Fujisawa, K.
 J. Gen. Virol. 73, 2725-2729, 1992
 A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated from a
 A:Reference number: JQ1925; MUID:93019030; PMID:1383400
 A:Accession: JQ1925
 A:Molecule type: RNA
 A:Residues: 1-520 <AB>
 A:Cross-references: DDBJ:U0687; NID:G221544; PID:BAA01529.1; PID:G221545
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: polypeptide; transmembrane protein
 F:1-191/Product: C protein #status predicted <CPR>
 F:192-383/Product: E1 protein #status predicted <E1>
 F:384-520/Product: E2/NS1 protein #status predicted <E2>

QY 1 MSTLPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLPRGPRLVGRATRTKTSERSOPRG 60
 DB 1 MSTNPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLTRGPRLVGRATRTKTSERSOPRG 60
 QY 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 DB 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPAS 191
 DB 181 LLSCLTTPAS 191
 QY 181 LLSCLTTPAS 191
 DB 181 LLSCLTTPAS 191

Query Match 91.1%; Score 952; DB 2; Length 520;
 Best Local Similarity 90.6%; Pred. No. 4.3e-73;
 Matches 173; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLPRGPRLVGRATRTKTSERSOPRG 60
 DB 1 MSTNPKPKRKNTNRRPPTDVKFGGGQIVGGVYLLTRGPRLVGRATRTKTSERSOPRG 60
 QY 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 DB 61 RROPPIKAPPOGRHMAOPGYWPLVGNBGCWAGWLLSPRGRPRMGPNDRRRSRNLG 120
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVDTLTGCFADLMGYIPVVGAPLVGAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPAS 191
 DB 181 LLSCLTTPAS 191
 QY 181 LLSCLTTPAS 191
 DB 181 LLSCLTTPAS 191

RESULT 22

JQ1926

polyprotein - hepatitis C virus (isolate HCV-476)

N/Contains: C protein; E1 protein; E2/NS1 protein

C/Species: hepatitis C virus

C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000

C/Accession: JQ1926

R/Abel, K.; Inchausti, G.; Fujisawa, K.

J. Gen. Virol. 73, 2725-2729, 1992

A/Title: Genomic characterization and mutation rate of hepatitis C virus isolated from a

A/Reference number: JQ1925; PMID:93019030; PMID:1383400

A/Accession: JQ1926

A/Molecule type: mRNA

A/Residues: 1-523 <ABE>

A/Cross-references: DDBJ: D10687

C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: polyprotein

F/1-191/Product: C protein #status predicted <CPR>

F/192-383/Product: E1 protein #status predicted <E1R>

F/523/Product: E2/NS1 protein #status predicted <E2P>

Query Match

Best Local Similarity 90.6%; Score 947; DB 2; Length 523;

Matches 172; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Db

1 MSTLPPKORTKNTNRPRPDVXPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

Db

1 MSTNPPKORTKNTNRPRPDVXPGGGQIVGGVYLLTRRGPRLGVATRTKTSERSQPRG 60

Qy

61 RROPPIKAROPGHRMAOPGYPMPLVGNCGAGWLLSPRGSRRPHMGPNDRPRRSRLNG 120

Db

61 RROPPIKARBERAMAOPGYPMPLVGNBGLGMAWLLSPRGSRRPHMGPNDRPRRSRLNG 120

Qy

121 KVITDTTCGFADLMGYIPVVGAPLVGAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180

Db

121 KVITDTTCGFADLMGYIPVVGAPLVGAALAHGVARIEDGINVATGNLPGCSFSIFLSA 180

Qy

181 LMSCLTTPASA 191

Db

181 LMSCLTTPASA 191

RESULT 23

S18032

genome polyprotein - hepatitis C virus (isolate JK4) (fragment)

N/Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C/Species: hepatitis C virus

C/Accession: S18032

R/Honda, M.; Kaneko, S.; Maseh, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A/Description: Sequence analysis of putative structural regions of Hepatitis C virus iso

A/Reference number: S18029

A/Accession: S18032

A/Molecule type: genomic RNA

A/Residues: 1-782 <HON>

A/Cross-references: EMBL: X61594

A/Experimental source: isolate JK4

C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F/1-191/Product: core protein #status predicted <MAT1>

F/192-383/Product: envelope protein 1 #status predicted <MAT2>

F/384-733/Product: NS1/E2 protein #status predicted <MAT3>

F/734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match

Best Local Similarity 90.6%; Score 947; DB 2; Length 782;

Matches 174; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy

1 MSTLPPKORTKNTNRPRPDVXPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

Db

1 MSTNPPKORTKNTNRPRPDVXPGGGQIVGGVYLLTRRGPRLGVATRTKTSERSQPRG 60

Qy

61 RROPPIKAROPGHRMAOPGYPMPLVGNCGAGWLLSPRGSRRPHMGPNDRPRRSRLNG 120

Db

61 RROPPIKARBERAMAOPGYPMPLVGNBGLGMAWLLSPRGSRRPHMGPNDRPRRSRLNG 120

Qy

121 KVITDTTCGFADLMGYIPVVGAPLVGAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180

Db

121 KVITDTTCGFADLMGYIPVVGAPLVGAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180

Qy

181 LMSCLTTPASA 190

Db

181 LMSCLTTPASA 190

RESULT 24

S41288

genome polyprotein - hepatitis C virus (fragment)

N/Contains: core protein; envelope protein; NS1 protein

C/Species: hepatitis C virus

C/Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C/Accession: S41288

R/Seeling, R.

submitted to the EMBL Data Library, December 1993

A/Reference number: S41288

A/Accession: S41288

A/Molecule type: genomic RNA

A/Residues: 1-492 <SEB>

A/Cross-references: EMBL: X76918

C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: capsid protein; core protein; envelope protein; nonstructural protein; polypr

F/1-191/Product: core protein #status predicted <COR>

F/192-372/Product: envelope protein #status predicted <ENV>

F/373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Query Match

Best Local Similarity 90.5%; Score 946; DB 2; Length 492;

Matches 172; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

Db

1 MSTLPPKORTKNTNRPRPDVXPGGGQIVGGVYLLPRRGPRLGVATRTKTSERSQPRG 60

Db

1 MSTNPPKORTKNTNRPRPDVXPGGGQIVGGVYLLTRRGPRLGVATRTKTSERSQPRG 60

Qy

61 RROPPIKAROPGHRMAOPGYPMPLVGNCGAGWLLSPRGSRRPHMGPNDRPRRSRLNG 120

Db

61 RROPPIKARBERAMAOPGYPMPLVGNBGLGMAWLLSPRGSRRPHMGPNDRPRRSRLNG 120

Qy

121 KVITDTTCGFADLMGYIPVVGAPLVGAALAHGVARIEDGINVATGNLPGCSFSIFLLA 180

Db

121 KVITDTTCGFADLMGYIPVVGAPLVGAALAHGVARIEDGINVATGNLPGCSFSIFLSA 180

Qy

181 LMSCLTTPASA 191

Db

181 LMSCLTTPASA 191

RESULT 25

PC2061

genome polyprotein N2 - hepatitis C virus

N/Contains: envelope protein E1; nonstructural protein E2/NS1

C/Species: hepatitis C virus

C/Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000

C/Accession: PC2061

R/Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.

Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994

A/Title: Identification of the third major genotype of hepatitis C virus in France.

A/Reference number: PC2060; PMID:94197744; PMID:8147893

A/Accession: PC2061

A/Molecule type: mRNA

A/Residues: 1-411 <LIJ>

A/Cross-references: GB: U12355; NID: g410169; PIDN: AAA20155.1; PID: g410170

C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein; i

F/192-383/Product: envelope protein E1 #status predicted <SBE>

F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NP>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 90.3%; Score 944; DB 2; Length 411;
Best Local Similarity 90.1%; Pred. No. 1,7e-72;
Matches 172; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RROPFKARPOGRRHMAQPGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNLG 120
DB 61 RROPFKARPSDRSVAQPGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNLG 120
QY 121 KVIDLTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LFSCLTHPAAS 191

RESULT 26

JQ0883
genome polypotein - hepatitis C virus (strain J7) (fragments)

N:Contains: NS5 protein

C:Species: hepatitis C virus

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Mar-2001

C:Accession: JQ0883

R:Okamoto, H.

submitted to JIPID, January 1991

A:Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis

A:Reference number: JQ0879

A:Accession: JQ0883

A:Molecule type: genomic RNA

A:Residues: 1-874 <OK>

A:Experimental source: strain J7

C:Superfamily: hepatitis C virus genome polypotein

C:Keywords: polypotein

F:510-874/Product: NS5 protein (fragment) #status predicted <NS5>

Query Match 90.0%; Score 941; DB 2; Length 874;
Best Local Similarity 89.5%; Pred. No. 6e-72;
Matches 171; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RROPFKARPOGRRHMAQPGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNLG 120
DB 61 RROPFKDRSRSTGSKGKGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNLG 120
QY 121 KVIDLTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 27

A44150
structural protein - hepatitis C virus

C:Species: hepatitis C virus

C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000

C:Accession: A44150

R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D.

Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992

A:Title: Interaction of immune sera with synthetic peptides corresponding to the structural

A:Reference number: A44150; MID:92228749; PMID:1373489
A:Accession: A44150
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: genomic RNA
A:Residues: 1-513 <CHI>
C:Superfamily: hepatitis C virus genome polypotein

Query Match 90.0%; Score 940; DB 2; Length 513;
Best Local Similarity 91.1%; Pred. No. 4,4e-72;
Matches 175; Conservative 5; Mismatches 10; Indels 2; Gaps 2;

QY 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RROPFKARPOGRRHMAQPGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNL 119
DB 61 RROPFKARPEGRTHAQPGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNL 119
QY 120 GKVIDTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLL 179
DB 120 GKVIDTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLL 179
QY 180 LLSCLTTPASA 191
DB 180 ALLSCLTPASA 191

RESULT 28

JQ0881
genome polypotein - hepatitis C virus (strain J6) (fragments)

N:Contains: NS5 protein

C:Species: hepatitis C virus

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 17-Nov-2000

C:Accession: JQ0881

R:Okamoto, H.

submitted to JIPID, January 1991

A:Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis

A:Reference number: JQ0879

A:Accession: JQ0881

A:Molecule type: genomic RNA

A:Residues: 1-874 <OK>

A:Experimental source: strain J6

C:Superfamily: hepatitis C virus genome polypotein

C:Keywords: polypotein

F:510-874/Product: NS5 protein (fragment) #status predicted <NS5>

Query Match 89.9%; Score 939; DB 2; Length 874;
Best Local Similarity 89.0%; Pred. No. 8,9e-72;
Matches 170; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
DB 1 MSTLPKPKQKTKNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RROPFKARPOGRRHMAQPGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNLG 120
DB 61 RROPFKDRSRSTGSKGKGYPMPLVGNBCCGAGMLSPRSGRPHMGPNDRRRSRNLG 120
QY 121 KVIDLTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTLTGGFADLMGYIPVVGAPLVGAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

RESULT 29

J01303

genome polypotein - hepatitis C virus (isolate HC-J6)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C/Species: hepatitis C virus
 C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
 C/Accession: J01303
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
 J. Gen. Virol. 72, 2697-2704, 1991
 A/Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
 A/Reference number: J01303; MWID:92044440; PMID:1658196
 A/Accession: J01303
 A/Molecule type: genomic RNA
 A/Residues: 1-3033 <OKA>
 A/Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651
 A/Experimental source: isolate HC-J6 from a Japanese individual
 C/Suprafamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1619-1619/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1321-1321/Region: nucleotide-binding motif B
 F:1620-1333/Region: DEHX motif
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 89.9%; Score 939; DB 1; Length 3033;

Best Local Similarity 89.0%; Pred. No. 2,9e-71;

Matches 170; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 1 MSTLPQRKTKNTNRPTDYKFPGGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 DB 1 MSTNPPQRKTKNTNRPPQDVKFPGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 QY 61 RROPTRKARPOGRRHNAOCPGYPMPLVGNCGAGWLLSPRGSRRPHMGPNDRRRSRNLG 120
 DB 61 RROPTRKDRSTGSKGKPGYPMPLVGNBGLGMAWLLSPRGSRRPHMGPNDRRRSRNLG 120
 QY 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTPASA 191
 DB 181 LLSCLTPASA 191

Job time : 22 secs

genome polyprotein - hepatitis C virus (strain HC-J6)
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HC-J6) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
 C/Species: hepatitis C virus
 C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 R/Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.;
 Virol. 188, 331-341, 1992
 A/Title: Full-length sequence of a hepatitis C virus genome having poor homology to repo
 A/Reference number: A40250; MWID:92203232; PMID:1314459
 A/Accession: A40250
 A/Molecule type: genomic RNA
 A/Residues: 1-3033 <OKA>
 A/Cross-references: GB:D010988; GB:D01221; NID:G221608; PIDN:BA01761.1; PID:G221609
 R/Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.
 J. Gen. Virol. 73, 1131-1141, 1992
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A/Reference number: PQ0393; MWID:92268871; PMID:1316939
 A/Accession: PQ0393
 A/Molecule type: genomic RNA
 A/Residues: 2678-2754 <CHA>
 A/Cross-references: DDBJ:D10134
 A/Experimental source: isolate E-D12

R/Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991
 A/Title: Distribution of pluvial HCV types in Japan.
 A/Reference number: PQ0554; MWID:92068204; PMID:1720309
 A/Accession: PQ0559
 A/Molecule type: mRNA
 A/Residues: 2678-2729 <KAT>
 A/Cross-references: GB:D10562; GB:D90518; NID:G221523; PIDN:BA01418.1; PID:G221524
 C/Suprafamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructural
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1234-1241/Region: nucleotide-binding motif A (P-loop)
 F:1316-1321/Region: nucleotide-binding motif B
 F:1320-1333/Region: DEHX motif
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 89.6%; Score 936; DB 1; Length 3033;

Best Local Similarity 88.5%; Pred. No. 5,2e-71;

Matches 169; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSTLPQRKTKNTNRPTDYKFPGGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 DB 1 MSTNPPQRKTKNTNRPPQDVKFPGGQIVGGVYLLPRGPRLGVRATRKTSERSQPRG 60
 QY 61 RROPTRKARPOGRRHNAOCPGYPMPLVGNCGAGWLLSPRGSRRPHMGPNDRRRSRNLG 120
 DB 61 RROPTRKDRSTGSKGKPGYPMPLVGNBGLGMAWLLSPRGSRRPHMGPNDRRRSRNLG 120
 QY 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 QY 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVDTLTGCFADIMGYIPVVGAPLGGVAAALAGVRAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTPASA 191
 DB 181 LLSCLTPASA 191

Search completed: May 12, 2004, 09:16:30
 Job time : 22 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2004, 09:08:40 ; Search time 18 Seconds
(without alignments)
552.522 Million cell updates/sec

Title: US-09-084-691b-206
Perfect score: 1045
Sequence: 1 MSTLPKPKRKTKNTRRPT.....CSFSIFLMLSLCTTPASA 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Search: 141681 seqs, 52070155 residues

number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 90 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	976	93.4	3011	1	P27958 h genome po
2	968	92.6	3011	1	POLG_HCV1
3	965	92.3	3010	1	POLG_HCVBK
4	965	92.3	3010	1	POLG_HCVJA
5	959	91.8	513	1	POLG_HCVJ2
6	956	91.5	3010	1	POLG_HCVJ2
7	952	91.1	520	1	POLG_HCVJK
8	948	90.7	3010	1	POLG_HCVJT
9	947	90.6	520	1	POLG_HCVH4
10	941	90.0	737	1	POLG_HCVJ7
11	939	89.9	3033	1	POLG_HCVJ6
12	938	89.8	737	1	POLG_HCVJ5
13	936	89.6	3033	1	POLG_HCVJ8
14	931	84.5	192	1	POLG_HCVJ1
15	931	84.5	321	1	POLG_HCVH8
16	359	34.4	321	1	POLG_HCVTH
17	298	28.5	309	1	POLG_HCVH7
18	97.5	9.3	420	1	LYOK_CHICK
19	93.5	8.9	256	1	PRIO_CAPI
20	93.5	8.9	256	1	PRIO_FELCA
21	93.5	8.9	256	1	PRIO_SHEEP
22	93.5	8.9	256	1	PRP2_BOVIN
23	93.5	8.9	256	1	PRIO_BOVIN
24	91.5	8.8	254	1	PRP1_TRAST
25	91	8.7	252	1	PRIO_RABIT
26	91	8.7	257	1	PRIO_PIG
27	90.5	8.7	256	1	PRIO_ODHEL
28	90.5	8.7	256	1	PRIO_ODHEL
29	90.5	8.7	256	1	PRP2_TRAST
30	88.5	8.5	252	1	V70K_TYMY
31	88	8.4	257	1	PRIO_CAJA
32	88	8.4	257	1	PRIO_MUSPF
33	88	8.4	257	1	PRIO_MUSVI

34	87.5	8.4	255	1	PRIO_CANPA	046501 canis fami1
35	86	8.2	437	1	NEOR_STRCY	P14501 streptomyce
36	85.5	8.2	252	1	PRIO_AREPA	P51446 ateles pan1
37	85.5	8.2	255	1	PRIO_SAIAC	P79141 camelus dro
38	85.5	8.2	259	1	PRIO_SAIAC	P40258 saimiri bcl
39	84	8.0	259	1	PRIO_TRIVU	P51780 trichosurus
40	83.5	8.0	232	1	PRIO_ATEGE	P40246 ateles geof
41	83.5	8.0	326	1	MIM1_CHICK	P08940 gallus gall
42	83.5	8.0	551	1	YO11_MYCTU	P71732 mycobacteri
43	83	7.9	241	1	PRIO_CALMO	P40248 callicebus
44	82.5	7.9	161	1	Y1A2_RHISP	P17964 thizobium s
45	82	7.8	201	1	SOBE_ONCVC	Q07449 onchocerca
46	82	7.8	628	1	V70K_TYMYC	P28478 turnip yell
47	82	7.8	750	1	ELS_CHICK	P07916 gallus gall
48	81.5	7.8	239	1	PRIO_AOTTR	P40245 actus triv1
49	81.5	7.8	415	1	SYN1_CANPA	Q62732 canis fami1
50	81	7.8	145	1	KIS1_HUMAN	Q15726 homo sapien
51	81	7.8	253	1	PRIO_GORGO	P40252 gorilla gor
52	81	7.8	253	1	PRIO_HUMAN	P04156 homo sapien
53	81	7.8	253	1	PRIO_PANTR	P40253 pan troglod
54	81	7.8	254	1	PRIO_PAT	P13852 ratius norv
55	80.5	7.7	254	1	PRIO_MOUSE	P04925 mus musculu
56	80	7.7	253	1	PRIO_PONPY	P40256 pongo pygma
57	80	7.7	254	1	PRIO_CRIGR	Q60506 cricetus
58	80	7.7	254	1	PRIO_SIGHI	Q92053 sigmodon hi
59	80	7.7	367	1	GCST_MYCLB	Q32955 mycobacteri
60	80	7.7	566	1	PKNE_MYCTU	P72001 mycobacteri
61	80	7.7	1097	1	S24C_ARATH	Q9M281 arabidopsis
62	79.5	7.6	648	1	GLT0_WHEAT	P10387 triticum ae
63	79.5	7.6	660	1	GLT3_WHEAT	P08488 triticum ae
64	79	7.6	342	1	FUT7_HUMAN	Q11130 homo sapien
65	78.5	7.5	241	1	PRIO_MANSP	P40255 mandrillus
66	78.5	7.5	254	1	PRIO_CRIMI	Q06468 cricetus
67	78.5	7.5	321	1	YB1_CHICK	Q06066 gallus gall
68	78	7.5	195	1	GRP3_DAUCA	P37705 daucus caro
69	77	7.4	254	1	PRIO_MESAV	P04273 mesocricetu
70	77	7.4	271	1	THTR_AZOVI	P52197 azotobacter
71	77	7.4	839	1	GLT5_WHEAT	P10388 triticum ae
72	76.5	7.3	324	1	YB1_HUMAN	P16991 h nuclease
73	76.5	7.3	325	1	NPDI_HUMAN	Q9N945 homo sapien
74	76.5	7.3	1069	1	S24B_ARATH	Q9M041 arabidopsis
75	76	7.3	252	1	DUT_HUMAN	P33316 homo sapien
76	76	7.3	382	1	METX_MYCLE	Q32874 mycobacteri
77	76	7.3	648	1	VP3_BPPH6	P11129 bacterioph
78	76	7.3	660	1	YH1_EBV	P01381 Epstein-bar
79	75.5	7.2	175	1	Y901_MYCTU	Q10559 mycobacteri
80	75.5	7.2	252	1	PRIO_CEBAP	P40249 cebus apell
81	75.5	7.2	322	1	YB1_MOUSE	P27817 m nuclease
82	75.5	7.2	593	1	Y917_MYCTU	Q05909 mycobacteri
83	75.5	7.2	697	1	SIRA_HUMAN	Q961c7 homo sapien
84	75.5	7.2	1219	1	SOB3_MOUSE	Q8V151 mus musculu
85	75	7.2	253	1	PRIO_COLGU	P40251 colobus gue
86	75	7.2	253	1	PRIO_PREFR	P40257 prebysyllis f
87	75	7.2	1324	1	IRSD_HUMAN	Q9Y412 homo sapien
88	74.5	7.1	625	1	DUS8_HUMAN	Q11202 homo sapien
89	74.5	7.1	825	1	SE5_RAT	Q63003 ratius norv
90	74.5	7.1	894	1	CYAS_HUMAN	Q95622 homo sapien

ALIGNMENTS

RESULT 1
ID POLG_HCVH STANDARD, PRT, 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepaticivirin)

DE (EC 3.4.21.99); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658600;
RA Inchauspé G., Zebbede S., Lee D.H.H., Sugtani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Messon T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RL "Structure of the hepatitis C virus RNA helicase domain.";
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493370;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100 (1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position. Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC -1- PM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, M67463; AAA45534.1; -.
DR PIR, A36814; GNNVCH.
DR PDB, 1HEI, 25-NOV-98.
DR PDB, 1A1V, 16-FEB-99.
DR PDB, 1A1R, 17-JUN-98.
DR MEROPS, S29.001; -.
DR MEROPS, U39.001; -.
DR TRASNSEC, T04155; -.
DR InterPro, IPR0009003; Cys_Ser_cryptin.
DR InterPro, IPR001410; DEAD.
DR InterPro, IPR0002522; HCV_core.
DR InterPro, IPR002521; HCV_core.
DR InterPro, IPR002519; HCV_env.
DR InterPro, IPR002531; HCV_nsi.

[illegible]

FT STRAND 1312 1316
 FT TURN 1317 1319
 FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT HELIX 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MM; 772CBB29CCD94753 CRC64;

Query Match 93.4%; Score 976; DB 1; Length 3011;
 Best Local Similarity 92.7%; Pred. No. 1.4e-74;
 Chex 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKKTKNTNRPRPDVYFPGGQI VGVYLLPRRGPRLGVRATRKTSERQPRG 60
 DB 1 MSTNPKPKKTKNTNRPRPDVYFPGGQI VGVYLLPRRGPRLGVRATRKTSERQPRG 60
 QY 61 RROPPIKARPOGRHMAOPGYPMPLTGNBCCGAGMLSPRGSRPHMGPNDRRRSRNIG 120
 DB 61 RROPPIKARPERBRTAOPGYPMPLTGNBCCGAGMLSPRGSRPHMGPNDRRRSRNIG 120
 QY 121 KYVIDLTTCGFADIMGYIPVAGAPLGGAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KYVIDLTTCGFADIMGYIPVAGAPLGGAAALAHGVRALIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTTPASA 191
 DB 181 LLSCLTTPASA 191

RESULT 2
 POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC F26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepaticin)
 DE (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M62321; AAA45676.1; -.
 CC PIR; A39166; GNWVC3.
 CC PDB; 1AIV; 16-FEB-98.
 CC PDB; 1HEI; 25-NOV-98.
 CC MEROPS; S29.001; -.
 CC DR MEROPS; U39.001; -.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_capsid.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4A.
 DR InterPro; IPR001490; HCV_NS4B.
 DR InterPro; IPR002868; HCV_NS5A.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02307; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.

DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KM Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KM Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 KM 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 92.6%; Score 968; DB 1; Length 3011;
 Best Local Similarity 91.6%; Pred. No. 6.5e-74;
 Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RN SEQUENCE FROM N.A.
 RN MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RA "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers";
 RL J. Virol. 65:1105-1113 (1991).
 RL [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RL phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618 (1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8661916;
 RA Love R.A., Parag H.E., Wickström J.A., Hostomsky Z., Habuka N.,
 RA Noonan E.W., Aach T., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RL trypsin-like fold and a structural zinc binding site";
 RL Cell 87:331-342 (1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RL virus: a 2.2-A resolution structure in a hexagonal crystal form";
 RL Protein Sci. 7:837-847 (1998).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipid-protein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M58335; AAA72945.1; -.
 DR ETR; A38465; GNMVTC.
 DR PDB; 1A1Q; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSU; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.

[illegible]

FT	STRAND	1050	1050
FT	STRAND	1050	1050
FT	STRAND	1059	1063
FT	STRAND	1068	1074
FT	TURN	1075	1076
FT	STRAND	1077	1081
FT	HELIX	1082	1085
FT	TURN	1086	1087
FT	STRAND	1090	1092
FT	TURN	1093	1094
FT	STRAND	1095	1097
FT	TURN	1101	1103
FT	TURN	1104	1107
FT	STRAND	1108	1112
FT	STRAND	1120	1120
FT	STRAND	1122	1122
FT	STRAND	1129	1133
FT	TURN	1135	1136
FT	STRAND	1139	1144
FT	STRAND	1149	1157
FT	HELIX	1158	1161
FT	TURN	1162	1163
FT	TURN	1165	1166
FT	STRAND	1168	1171
FT	TURN	1172	1174
FT	STRAND	1175	1186
FT	TURN	1187	1188
FT	STRAND	1189	1197
FT	HELIX	1198	1202
FT	TURN	1203	1204
FT	STRAND	1680	1689
SO	SEQUENCE	3010 AA;	327189 MM; F8422D5ECCFPD9C CRC64;
Query Match			
Best Local Similarity 92.1%; Score 965; DB 1; Length 3010;			
Matches 176; Conservative 5; Mismatches 10; Indels 0; Gaps 0;			
Qy	1	MSLTPKQRTKNTNRPDVKFPGGGQIVGVYLLPRGPRLGVAATKRTSERSQRC	60
Db	1	MSLTPKQRTKNTNRPDVKFPGGGQIVGVYLLPRGPRLGVAATKRTSERSQRC	60
Qy	61	RROPFPARPOGCHMAOPGYPMPLYNEGCGMAGMLSPGSRPHMGPNDRRRSRILG	120
Db	61	RROPFPARPOGCHMAOPGYPMPLYNEGCGMAGMLSPGSRPHMGPNDRRRSRILG	120
Qy	121	KVIDTLTCGFADLMGYIPVVGAPLGVAAALAHGVARIEDGINVATGNTPCSCFSIFLLA	180
Db	121	KVIDTLTCGFADLMGYIPVVGAPLGVAAALAHGVARIEDGINVATGNTPCSCFSIFLLA	180
Qy	181	LLSCLTTPASA 191	
Db	181	LLSCLTTPASA 191	
RESULT 4			
POLG_HCVJA STANDARD; PRT; 3010 AA.			
AC	P26662;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DB	Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);		
DB	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2		
DB	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)		
DB	(EC 3.4.22.-); Processase/helicase NS3 (P70) (Hepactivin)		
DB	(EC 3.4.21.99); Nonstructural protein NS4A (P4); Nonstructural protein		
DB	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein		
DB	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]		
OS	Hepatitis C virus (Isolate Japanese) (HCV).		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
CC	Hepacivirus.		
OX	NCBI_Taxid=11116;		
XP	[1		
FP	SEQUENCE FROM N.A.		

RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Iizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10074; BAA00968.1; -;
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Prodom; PD186062; HCV_NS1; 1.
 DR Polyprotein. Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1
 FT 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >513
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT NON_TER 513 513
 SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;
 Query Match 91.8%; Score 959; DB 1; Length 513;
 Best Local Similarity 91.6%; Pred. No. 6,3e-74;
 Matches 175; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

RESULT 6
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/Helicase NS3 (P70) (Hepatitisin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (Isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' terminus of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M84754; -; NOT_ANNOTATED_CDS.
 DR PIR; A40244; GNMVW.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; 1N63; 08-APR-98.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4.
 DR InterPro; IPR001490; HCV_NS4B.
 DR InterPro; IPR002868; HCV_NS5A.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KM Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 KM 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 91.5%; Score 956; DB 1; Length 3010;
 Best Local Similarity 91.6%; Pred. No. 6,7e-73;
 Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;
 Oy 1 MSTLPKPKTKKNTNRPPDVKFPGGGQIVGGYLLPRGPRLCVATKTSERSQPRG 60
 Db 1 MSTNGKPKTKKNTNRPPDVKFPGGGQIVGGYLLPRGPRLCVATKTSERSQPRG 60
 Oy 61 RRGPPIKAROPGGRHMAQPGYPMPLVNGEGCGMAGLISRGSPHMGPDPRRSNGLG 120
 Db 61 RRGPPIKAROPGGRHMAQPGYPMPLVNGEGCGMAGLISRGSPHMGPDPRRSNGLG 120
 Oy 121 KVIDITLCGFADLMGYIPVVGAPLGVAAALAGVRAIEDGIVATGNLPGCSFSIFLLA 180
 Db 121 KVIDITLCGFADLMGYIPVVGAPLGVAAALAGVRAIEDGIVATGNLPGCSFSIFLLA 180
 Oy 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191

RESULT 7
 POLG_HCVHK STANDARD; PRT; 520 AA.
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polypeptide [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1) (Fragment).
 OS Hepatitis C virus (isolate HCV-KF) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxId=31644;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchausti G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan."
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -i- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10687; BAA01529.1; -
 DR PIR; J01925; J01925.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KM Polypeptide; Glycoprotein; Coat protein; Envelope protein;
 KM Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;
 Query Match 91.1%; Score 952; DB 1; Length 520;
 Best Local Similarity 90.6%; Pred. No. 2.5e-73;
 Matches 173; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTLPEKPKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGLGVATRTKTSERSQPRG 60
 DB 1 MSTNPKPKTKNTNRPTDVKFPGGGQIVGGVYLLTRRGLGVATRTKTSERSQPRG 60
 QY 61 RROPPEKARQPGGRHMAQPGYPMPLVYNGECGNAWMLSPRGRPHMGPNDRSRNLTG 120
 DB 61 RROPPEKARQPGGRHMAQPGYPMPLVYNGECGNAWMLSPRGRPHMGPNDRSRNLTG 120
 QY 121 KVIDTLTTCGADLMGYTPVVGAPLGGVAAALAHGVRALENGIVATGNLPGCSFSTFLA 180
 DB 121 KVIDTLTTCGADLMGYTPVVGAPLGGVAAALAHGVRALENGIVATGNLPGCSFSTFLA 180
 QY 181 LMSCLTTPASA 191
 DB 181 LMSCLTTPASA 191
 RESULT 8
 POLG HCVUT STANDARD; PRT: 3010 AA.
 A 000269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (BC 3.4.21.98); Protease/helicase NS3 (P70) (Hepatitisin)
 DE (BC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (BC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-7) (HCV) (HCV)
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=31642;
 RX MEDLINE=92295714; Pubmed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals";
 RL Virus Res. 23:39-53(1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PERTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D11168; BAA01943.1; -
 CC PIR: A45573; A45573.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC InterPro: IPR005003; Cys_Ser_lysine.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.

DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4A.
 DR InterPro: IPR001490; HCV NS4B.
 DR InterPro: IPR002868; HCV NS5A.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01539; HCV core; 1.
 DR Pfam: PF01560; HCV env; 1.
 DR Pfam: PF01538; HCV NS1; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4A; 1.
 DR Pfam: PF01001; HCV NS4B; 1.
 DR Pfam: PF01506; HCV NS5A; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEKDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 192 191
 FT CHAIN 384 383
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 90.7%; Score 948; DB 1; Length 3010;
 Best Local Similarity 90.6%; Pred. No. 3; 2e-72;
 Matches 173; Conservative 6; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MSTLPEKPKTKNTNRPTDVKFPGGGQIVGGVYLLPRRGLGVATRTKTSERSQPRG 60

DB 1 MSTNPKRQKTKNTYRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRAIRKTSERQPRG 60
 QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNCGMAGMLSPRSRPHMGNDPRRRSRNLG 120
 DB 61 RROPPIKARPEGRMAOPGYPMPLVGNCGMAGMLSPRSRPHMGNDPRRRSRNLG 120
 QY 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LMSCLTTPASA 191
 DB 181 LMSCLTTPASA 191

RESULT 9
 POLG_HCVH4 STANDARD; PRT; 520 AA.
 AC 001404;
 DB 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polypeptide [contains: Capsid protein C (core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HCV-476) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis C virus.
 NC NCB1_TaxID=31643;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchausti G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729 (1992).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DB EMBL: D10688; BAA01530.1; -
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR ProDom: PD16062; HCV_NSI; 1.
 DR PolyProtein: Glycoprotein; Coat protein; Envelope protein;
 DR Transmembrane; Nonstructural protein.
 KW INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305

FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;
 Query Match 90.6%; Score 947; DB 1; Length 520;
 Best Local Similarity 90.1%; Pred. No. 6.6e-73;
 Matches 172; Conservative 7; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MSTNPKRQKTKNTYRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRAIRKTSERQPRG 60
 DB 1 MSTNPKRQKTKNTYRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRAIRKTSERQPRG 60
 QY 61 RROPPIKARPOGRHMAOPGYPMPLVGNCGMAGMLSPRSRPHMGNDPRRRSRNLG 120
 DB 61 RROPPIKARPEGRMAOPGYPMPLVGNCGMAGMLSPRSRPHMGNDPRRRSRNLG 120
 QY 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 DB 121 KVIDLTCGFPADLMGYIPVVGAPLGGVAALAHGVAIEDGINVATGNLPGCSFSIFLLA 180
 QY 181 LMSCLTTPASA 191
 DB 181 LMSCLTTPASA 191

RESULT 10
 POLG_HCVU7 STANDARD; PRT; 737 AA.
 ID POLG_HCVU7
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polypeptide [contains: Capsid protein C (core protein); Matrix
 DE protein (NS1 and NS2) (Fragment).
 DE Hepatitis C virus (isolate HC-47) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis C virus.
 NC NCB1_TaxID=11114;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9220232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Iizuka H.,
 RA Tanaka T., Fukuda S., Tenda F., Mochi S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341 (1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DB EMBL: D10077; BAA00971.1; -
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.


```

FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 365 POTENTIAL.
FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1234 1241 ATP (POTENTIAL).
FT SITE 1320 1323 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 329165 MW; P957FSC1A273BE9E CRC64;

```

```

Query Match 89.9%; Score 939; DB 1; Length 3033;
Best Local Similarity 89.0%; Pred. No. 1,8e-71;
Matches 170; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 1 MSTLPRPQRTKNTNRRPTDYKFPFGGQIVGVYLLPRGPRLGVRATRKTSERQPRG 60
DB 1 MSTNPRPQRTKNTNRRPQDYKFPFGGQIVGVYLLPRGPRLGVRATRKTSERQPRG 60
QY 61 RROPPIKAROPGGRHNAQPGYPMPLVGNBSCGAGWLLSPRGSRPFGMPNDPRRSRNIG 120
DB 61 RROPPIKDRSRSTKSKNGKPGYPMPLVGNBGLMAGWLLSPRGSRPFGMPNDPRHRSRNVG 120
QY 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

```

```

RESULT 12
POLG_HCVUS STANDARD; PRT; 737 AA.
AC P27960;

```

```

DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide (Contains: Capsid protein C (core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE Hepatitis C virus (isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitis C virus.
OC NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; Pubmed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Iizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mielhro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."

```

```

RL Virology 186:331-341(1992).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipid-protein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
CC EMBL: D10075; BAA00969.1; -.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC INIT_MET 1
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 733
CC CHAIN 734 >737
CC TRANSMEM 347 365
CC CARBOHYD 196 196
CC CARBOHYD 209 209
CC CARBOHYD 234 234
CC CARBOHYD 305 305
CC CARBOHYD 417 417
CC CARBOHYD 423 423
CC CARBOHYD 430 430
CC CARBOHYD 448 448
CC CARBOHYD 477 477
CC CARBOHYD 534 534
CC CARBOHYD 542 542
CC CARBOHYD 558 558
CC CARBOHYD 578 578
CC CARBOHYD 627 627
CC CARBOHYD 649 649
CC NON_TER 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

```

```

Query Match 89.8%; Score 938; DB 1; Length 737;
Best Local Similarity 89.0%; Pred. No. 5,4e-72;
Matches 170; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 1 MSTLPRPQRTKNTNRRPTDYKFPFGGQIVGVYLLPRGPRLGVRATRKTSERQPRG 60
DB 1 MSTNPRPQRTKNTNRRPQDYKFPFGGQIVGVYLLPRGPRLGVRATRKTSERQPRG 60
QY 61 RROPPIKAROPGGRHNAQPGYPMPLVGNBSCGAGWLLSPRGSRPFGMPNDPRRSRNIG 120
DB 61 RROPPIKDRSRSTKSKNGKPGYPMPLVGNBGLMAGWLLSPRGSRPFGMPNDPRHRSRNVG 120
QY 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAALIEDGINVATGNLPGCSFSIFLLA 180
DB 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAALIEDGINVATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
DB 181 LLSCLTTPASA 191

```


DB 181 LSCITVPVSA 191

RESULT 13

POLG_HCVU8 STANDARD; PRT; 3033 AA.

AC P2661;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last annotation update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P22); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

DE Hepatitis C virus (isolate HC-J8) (HCV).

OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

CC NCBI_Taxid=11115;

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92230232; PubMed=1314459;

RA Okamoto H., Kuzai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T., Fukuda S., Tsuda F., Mishihiro S.;

RT "Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes."

RT Virology 188:331-341(1992).

CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).

CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

CC -----

DR EMBL; D10988; BAA01761.1; -

DR PIR; A40250; GNMVJ8.

DR HSSP; P27958; 1HE1.

DR MEROPS; S29.001; -

DR MEROPS; U39.001; -

DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002531; HCV_NS1.

DR InterPro; IPR002518; HCV_NS2.

DR InterPro; IPR000745; HCV_NS4A.

DR InterPro; IPR001490; HCV_NS4B.

DR InterPro; IPR002868; HCV_NS5A.

DR InterPro; IPR002166; HCV_RdRp.

DR InterPro; IPR004109; Peptidase_C39.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NS1; 1.

DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam; PF02907; HCV_NS3; 1.

DR Pfam; PF01006; HCV_NS4A; 1.

DR Pfam; PF01001; HCV_NS4B; 1.

DR Pfam; PF01506; HCV_NS5A; 1.

DR Pfam; PF00998; Viral_RdRp; 1.

DR Pfam; PF016062; HCV_NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural

INT_MET 1 1

FT CHAIN 1 115

FT CHAIN 116 191

FT CHAIN 192 383

FT CHAIN 384 733

FT CHAIN 734 1010

FT CHAIN 1011 1619

FT CHAIN 1620 1866

FT CHAIN 1867 2017

FT CHAIN 2018 3033

FT TRANSMEM 347 369

FT ACT_SITE 1087 1087

FT ACT_SITE 1111 1111

FT ACT_SITE 1165 1165

FT NP_BIND 1234 1241

FT SITE 1320 1323

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 233 233

FT CARBOHYD 299 299

FT CARBOHYD 305 305

FT CARBOHYD 417 417

FT CARBOHYD 423 423

FT CARBOHYD 430 430

FT CARBOHYD 448 448

FT CARBOHYD 477 477

FT CARBOHYD 534 534

FT CARBOHYD 542 542

FT CARBOHYD 558 558

FT CARBOHYD 578 578

FT CARBOHYD 627 627

FT CARBOHYD 649 649

FT CARBOHYD 1091 1091

FT CARBOHYD 2038 2038

FT CARBOHYD 2359 2359

FT CARBOHYD 2811 2811

SO SEQUENCE 3033 AA; 330177 MW; 1A17387E3381FDDA CRC64;

Query Match 89.6%; Score 936; DB 1; Length 3033;

Best Local Similarity 88.5%; Pred. No. 3.3e-71;

Matches 169; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRLYATRTKTSRSQPRG 60

DB 1 MSTNPKPKRKTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRLYATRTKTSRSQPRG 60

QY 61 RROPPIKARPOGSHNAOPGYPMPLVYNGECGAGMLSPRGSRPHWPNDRPRRSNTLG 120

DB 61 RROPPIKARPOGSHNAOPGYPMPLVYNGECGAGMLSPRGSRPHWPNDRPRRSNTLG 120

QY 121 KVIDLTGCGADLMGTYIPVVGARPLGVAAALAGVRAIEGIVYATGNLPGCSFSTFLA 180

DB 121 RVIDITTCGADLMGTYIPVVGARPLGVAAALAGVRAIEGIVYATGNLPGCSFSTFLA 180

QY 181 LSCITVPVSA 191

DB 181 LSCITVPVSA 191

```

RESULT 14
POLG_HCV_E1 STANDARD; PRT; 192 AA.
AC P27954;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE Major envelope protein E (Fragment).
DE Genome polyprotein [contains: Matrix protein (Envelope protein M);
DE Hepatitis C virus (isolate HCV) (HCV).
DE Hepatitis C virus (isolate HCV) (HCV).
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11107;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RL Virology 180:842-848(1991).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X51313; CAAT7295.1; -.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane.
FT NON TER 1
FT CHAIN 75
FT CARBOHYD 80
FT CARBOHYD 93
FT CARBOHYD 118
FT CARBOHYD 189
FT CARBOHYD 192
FT NON TER 192
SQ SEQUENCE 192 AA; 20315 MW; 98E48F4C335A86C CRC64;

Query Match 34.5%; Score 361; DB 1; Length 192;
Best Local Similarity 90.7%; Pred. No. 8.8e-24;
Matches 68; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 117 RNIGKVIDLTGCGFADLMGYIPVVGAPLGVAALAHGVAIEEDGINVATGNLPGCSFSI 176
DB 1 RNIGKVIDLTGCGFADLMGYIPVVGAPLGVAALAHGVAIEEDGINVATGNLPGCSFSI 60

QY 177 FLTALISCLTTPASA 191
DB 61 FLTALISCLTTPASA 75

RESULT 15
POLG_HCV_H8 STANDARD; PRT; 321 AA.
AC P27956;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Genome polyprotein [contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1 (Fragment).
DE Hepatitis C virus (isolate HCT18) (HCV).
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RL Virology 180:842-848(1991).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X51313; CAAT7291.1; -.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON TER 1
FT CHAIN 76
FT CHAIN 267
FT CARBOHYD 268
FT CARBOHYD 80
FT CARBOHYD 93
FT CARBOHYD 118
FT CARBOHYD 189
FT CARBOHYD 301
FT CARBOHYD 307
FT CARBOHYD 314
FT CARBOHYD 321
FT NON TER 321
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;

Query Match 34.5%; Score 361; DB 1; Length 321;
Best Local Similarity 90.7%; Pred. No. 1.5e-23;
Matches 68; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 117 RNIGKVIDLTGCGFADLMGYIPVVGAPLGVAALAHGVAIEEDGINVATGNLPGCSFSI 176
DB 1 RNIGKVIDLTGCGFADLMGYIPVVGAPLGVAALAHGVAIEEDGINVATGNLPGCSFSI 60

QY 177 FLTALISCLTTPASA 191
DB 61 FLTALISCLTTPASA 75

RESULT 16
POLG_HCV_H8 STANDARD; PRT; 321 AA.
AC P27957;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins."
RL Virology 180:842-848(1991).
CC -1 SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC 110protein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X53134; CAA37294.1; -
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC
CC FT CHAIN 1 1
CC FT NON_TER 1 1
CC FT CHAIN 76 75
CC FT CHAIN 267 267
CC FT CHAIN 268 >321
CC FT CARBOHYD 80 80
CC FT CARBOHYD 93 93
CC FT CARBOHYD 118 118
CC FT CARBOHYD 189 189
CC FT CARBOHYD 301 301
CC FT CARBOHYD 307 307
CC FT CARBOHYD 314 314
CC FT NON_TER 321 321
CC SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;
Query Match 34.4%; Score 359; DB 1; Length 321;
Best Local Similarity 89.3%; Pred. No. 2,2e-23;
Matches 67; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 117 RNIGKVIDITFCGADLMGYIPVVGAPLGVAAALAHGVRALIEDGINVATGNLPGCSFSI 176
DB 1 RNIGKVIDITFCGADLMGYIPVVGAPLGVAAALAHGVRALIEDGINVATGNLPGCSFSI 60
QY 177 FILALLSCLTTPASA 191
DB 61 FILALLSCLTTPASA 75
RESULT 17
POLG_HCVH7 STANDARD; PRT; 309 AA.
AC F27955;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT27) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11109;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins."
RL Virology 180:842-848(1991).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X53133; CAA37293.1; -
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC
CC FT CHAIN 1 1
CC FT NON_TER 1 1
CC FT CHAIN 64 63
CC FT CHAIN 256 >309
CC FT CARBOHYD 68 68
CC FT CARBOHYD 81 81
CC FT CARBOHYD 106 106
CC FT CARBOHYD 177 177
CC FT CARBOHYD 289 289
CC FT CARBOHYD 295 295
CC FT CARBOHYD 302 302
CC FT NON_TER 309 309
CC SQ SEQUENCE 309 AA; 32922 MW; 6E858E9C3D0B9EA9 CRC64;
Query Match 28.5%; Score 298; DB 1; Length 309;
Best Local Similarity 88.9%; Pred. No. 3e-18;
Matches 56; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 129 GFADLMGYIPVVGAPLGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLALLSCLTTP 188
DB 1 GFADLMGYIPVVGAPLGVAAALAHGVRALIEDGINVATGNLPGCSFSIFLALLSCLTTP 60
QY 189 ASA 191
DB 61 ASA 63
RESULT 18
LYOX_CHICK STANDARD; PRT; 420 AA.
AC Q05063;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-lysine 6-oxidase precursor (EC 1.4.3.13) (lysoyl oxidase).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

CC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=93077497; PubMed=1360009;
 RA Wu Y., Rich C.B., Lincicum J., Trackman P.C., Kagan H.M., Foster J.A.;
 RT "Characterization and developmental expression of chick aortic lysyl
 oxidase";
 RL J. Biol. Chem. 267:24199-24206(1992).
 CC -1- FUNCTION: Responsible for the posttranslational oxidative
 cleavage of peptidyl lysine residues in precursors to fibrous
 collagen and elastin.
 CC -1- FUNCTION: IN ADDITION TO CROSS LINKING OF EXTRACELLULAR MATRIX
 PROTEINS, IT MAY HAVE A DIRECT ROLE IN TUMOR SUPPRESSION.
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysyl-peptide + H(2)O + O(2) =
 peptidyl-L-allyl-peptide + NH(3) + H(2)O(2).
 CC -1- COFACTOR: Copper and LTO (By similarity).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- DEVELOPMENTAL STAGE: INCREASES BETWEEN DAY 8 AND 16 OF EMBRYONIC
 DEVELOPMENT, DURING AORTIC EMBRYOGENESIS, IN DIRECT PROPORTION
 TO TOTAL PROTEIN SYNTHESIS.
 CC -1- PTM: The lysine tyrosylquinone cross-link (LTO) is generated by
 condensation of the epsilon-amino group of a lysine with a
 topaquinone produced by oxidation of tyrosine.
 CC -1- SIMILARITY: Belongs to the lysyl oxidase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M97881; AAA48942.1; -.
 DR PIR; A45166; A45166.
 DR InterPro; IPR001695; Lysyl oxidase.
 DR Pfam; PF01186; Lysyl oxidase; 1.
 DR PRINTS; PR00074; LYSTOXIDASE.
 DR PRODEM; PD013867; Lysyl oxidase; 1.
 DR PROSITE; PS00926; LYSTOXIDASE; 1.
 DR Oxidoreductase; Copper; Signal; Glycoprotein; LTO.
 KW SIGNAL 1 21
 FT PROPEP 1 22 153
 FT CHAIN 154 420
 FT DOMAIN 216 420
 FT METAL 295 295
 FT METAL 297 297
 FT METAL 299 299
 FT CROSSLINK 323 356
 FT MOD RES 358 358
 FT CARBOHYD 78 78
 SQ SEQUENCE 420 AA; 48152 MW; E0CB81DD625F5C2 CRC64;
 Query Match 9.3%; Score 97.5; DB 1; Length 420;
 Best Local Similarity 28.2%; Pred. No. 0.38; Indels 49; Gaps 7;
 Matches 35; Conservative 10; Mismatches 30;
 QY 22 VFFPGGQIVT---GGVLLPRRG-----PRLGVAATKTSERSQ 57
 Db 38 IQWENNGQYVSYLSQGAQVPPRRRQGAEPASSPVLLRNGSVPRRAAAAARPPQEPQ 97
 QY 58 PGRGRQPIPYA--RQPG--RHMAQPGYPWDLVNEGGAGWILLSPRGS--RPHWGPNDR 113
 Db 98 PQAPQPPRRSRSPRLGRHWFQAGY-----RAPSGRAP-----APR 136
 QY 114 RRSR 117
 Db 137 RRPR 140

RESULT 19
 ID PRIOR_CAPIH STANDARD; PRT; 256 AA.
 AC P52113;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Major prion protein precursor (Prp).
 GN PRNP OR PRP.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OX NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Anglo-Nubian; TISSUE=Peripheral blood lymphocytes;
 RA Martin T.C., Hughes S.L., Hughes K.J., Dawson M.;
 RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT MET-142.
 RC STRAIN=Anglo-Nubian;
 RX MEDLINE=97081203; PubMed=8922485;
 RA Goldmann W., Martin T., Foster J., Hughes S., Smith G., Hughes K.,
 Dawson M., Hunter N.;
 RT "Novel polymorphisms in the caprine Prp gene: a codon 142 mutation
 associated with scrapie incubation period.";
 RL J. Gen. Virol. 77:2885-2891(1996).
 RN [3]
 RP ERRATUM.
 RA Goldmann W., Martin T., Foster J., Hughes S., Smith G., Hughes K.,
 Dawson M., Hunter N.;
 RL J. Gen. Virol. 78:697-697(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=African dwarf; TISSUE=Blood;
 RX MEDLINE=96356540; PubMed=8746958;
 RA Obernauer G., Kretzschmar H.A., Hafner A., Heuback D., Dahme E.;
 RT "Spongiform central nervous system myelinopathy in African dwarf
 goats";
 RL J. Comp. Pathol. 113:357-372(1995).
 CC -1- FUNCTION: The function of Prp is not known. Prp is encoded in the
 host genome and is expressed both in normal and infected cells.
 CC -1- SUBUNIT: Prp has a tendency to aggregate yielding polymers called
 "rods".
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- DISEASE: POLYOMORPHISM AT POSITION 171 MAY BE RELATED TO THE
 ALLELES OF SCRAPIE INCUBATION-CONTROL (SIC) GENE IN THIS SPECIES.
 CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS
 INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,
 CREUTZFELDT-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME
 (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
 TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -1- SIMILARITY: Belongs to the prion family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X74758; CAA52774.1; -.
 CC EMBL; X91999; CAA63050.1; -.
 CC EMBL; S82626; AAD14409.1; -.
 DR PIR; S37149; S37149.
 DR HSSP; P10279; 1DWY.
 DR InterPro; IPR000817; Prion.
 DR Pfam; PF00377; prion; 1.
 DR Pfam; PF03991; prion_octapep; 5.
 DR PRINTS; PR00341; PRION.

DR SMART; SM00157; PRP; 1.
 DR PROSITE; PS00291; PRION 1; 1.
 DR PROSITE; PS00706; PRION 2; 1.
 KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.
 FT CHAIN 1 24 BY SIMILARITY.
 FT CHAIN 2 25 MAJOR PRION PROTEIN.
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 182 217 BY SIMILARITY.
 FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-
 FT REPEAT 54 62 0.
 FT REPEAT 63 70 1.
 FT REPEAT 71 78 2.
 FT REPEAT 79 86 3.
 FT REPEAT 87 95 4.
 FT VARIANT 142 142 5.
 FT VARIANT 142 142 I -> M (APPEARS TO BE ASSOCIATED WITH
 DIFFERING DISEASE INCUBATION PERIODS IN
 GOATS EXPERIMENTALLY INFECTED WITH
 ISOLATES OF BOVINE SPONGIFORM
 ENCEPHALOPATHY OR SHEEP SCRAPIE).
 SQ SEQUENCE 256 AA; 27897 MW; BD45795F6FD9746 CRC64;
 Query Match 8.9%; Score 93.5; DB 1; Length 256;
 Best Local Similarity 29.2%; Pred. No. 0.5;
 Matches 47; Conservative 6; Mismatches 57; Indels 51; Gaps 12;
 QY 17 RRPDYKPPGGQIVGVYLLPRGRPLGVRAIRKTSRSQPRGRPIPKA---RQPG 72
 DB 26 KRKP-----PGGWNNTG-----SRYPQG-----SPGGRNP-FQGGGSGWQPH 64
 QY 73 GRWMAQ---GYWPLVYNGCGM-----AGWLLSPGRSPHMGPRRRRNIGKVI 123
 DB 65 GGGMGPPHGGGQGP-----HGGMGQPHGGGGM--GGGSHSQW--NKPSKPTNMKHYA 116
 QY 124 DLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINY 164
 DB 117 GAAAG-----AVVGG-LGGYMGSMRPLIHFGNDY 148
 RESULT 20
 Prio FELCA STANDARD; PRT; 256 AA.
 ID 018754; 019016;
 AC 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 PR Major prion protein precursor (Prp).
 OS PRNP OR PRP.
 OC Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_TaxID=9685;
 RN NCBI_TaxID=9685;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood, and Brain;
 RA Rohrer R.G., Edelman D., Protzman J.L.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 112-235 FROM N.A.
 RA Taylor M.S., Newton D.J., Flanagan B.F., Christas S.B.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: The function of Prp is not known. Prp is encoded in the
 host genome and is expressed both in normal and infected cells.
 CC -1- SUBUNIT: Prp has a tendency to aggregate yielding polymers called
 "rods".
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS
 INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,
 CREUTZFELDT-JAKOB DISEASE (CJD), GERSTMAN-STRAUSSLER SYNDROME
 (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
 TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -1- SIMILARITY: Belongs to the prion family.

CC -1- CAUTION: Ref.1 sequence seems to be incorrect, it is too close in
 CC sequence to that of sheep to be taxonomically correct. We have
 CC used Ref.2 sequence in the region where it is available (112-235),
 CC but the rest of the sequence probably contains incorrect residues.
 CC -1- DATABASE: NAME=Cat Prp; NOTE=Web page on cat sequence problems;
 CC WWW="http://www.mad-cow.org/~tom/cat/prion.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL; AF003087; AAB70468.1; -.
 DR EMBL; Y13698; CAA74032.1; -.
 DR HSSP; P04925; IAG2.
 DR InterPro; IPR000817; Prion.
 DR Pfam; PF00377; prion; 1.
 DR Pfam; PF03991; prion, octapep; 5.
 DR PRINTS; PR00341; PRION.
 DR SMART; SM00157; PRP; 1.
 DR PROSITE; PS00291; PRION 1; 1.
 DR PROSITE; PS00706; PRION 2; 1.
 KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.
 FT CHAIN 1 24 BY SIMILARITY.
 FT CHAIN 2 25 MAJOR PRION PROTEIN.
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 182 217 BY SIMILARITY.
 FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-
 FT REPEAT 54 62 0.
 FT REPEAT 63 70 1.
 FT REPEAT 71 78 2.
 FT REPEAT 79 86 3.
 FT REPEAT 87 95 4.
 FT REPEAT 162 162 5.
 FT CONFLICT 162 162 D -> N (IN REF. 1).
 FT CONFLICT 180 180 R -> H (IN REF. 1).
 FT CONFLICT 190 190 R -> H (IN REF. 1).
 FT CONFLICT 206 206 M -> I (IN REF. 1).
 FT CONFLICT 218 218 V -> I (IN REF. 1).
 FT CONFLICT 223 223 K -> R (IN REF. 1).
 FT CONFLICT 232 232 R -> G (IN REF. 1).
 FT CONFLICT 235 235 A -> V (IN REF. 1).
 SQ SEQUENCE 256 AA; 27975 MW; 7C687C3BC6BBB9 CRC64;
 Query Match 8.9%; Score 93.5; DB 1; Length 256;
 Best Local Similarity 29.2%; Pred. No. 0.5;
 Matches 47; Conservative 6; Mismatches 57; Indels 51; Gaps 12;
 QY 17 RRPDYKPPGGQIVGVYLLPRGRPLGVRAIRKTSRSQPRGRPIPKA---RQPG 72
 DB 26 KRKP-----PGGWNNTG-----SRYPQG-----SPGGRNP-FQGGGSGWQPH 64
 QY 73 GRWMAQ---GYWPLVYNGCGM-----AGWLLSPGRSPHMGPRRRRNIGKVI 123
 DB 65 GGGMGPPHGGGQGP-----HGGMGQPHGGGGM--GGGSHSQW--NKPSKPTNMKHYA 116
 QY 124 DLTTCGFADLMGYIPVVGAPLGGVAAALAHGVAIEDGINY 164
 DB 117 GAAAG-----AVVGG-LGGYMGSMRPLIHFGNDY 148
 RESULT 21
 Prio SHEEP STANDARD; PRT; 256 AA.
 ID P23907;
 AC 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 PR Major prion protein precursor (Prp).

PRP2_BOVIN STANDARD; PRT; 256 AA.

ID PRP2_BOVIN STANDARD; PRT; 256 AA.

AC 001880;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Major prion protein 2 precursor (Prp) (Major scrapie-associated fibril protein 2).

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bovinae; Bos.

OC NCBI_TaxID=9913;

OX NCBI_TaxID=9913;

RA MEDLINE=93118243; PubMed=1362024;

RA Yoshimoto J., Iinuma T., Ishiguro N., Horiuchi M., Imanura M., Shingawa M.;

RA "Comparative sequence analysis and expression of bovine Prp gene in mouse L-929 cells."

RL Virus Genes 6:343-356(1992).

CC -1- FUNCTION: The function of Prp is not known. Prp is encoded in the host genome and is expressed both in normal and infected cells.

CC -1- SUBUNIT: Prp has a tendency to aggregate yielding polymers called "rods".

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.

CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU, CREDITFIELD-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE), TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.

CC -1- SIMILARITY: Belongs to the prion family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: J01614; BAA01469.1; -

DR PIR: J02668; J02668.

DR HSSP: P10279; JDMY.

DR InterPro: IPR001610; PAC.

DR InterPro: IPR000817; Prion.

DR Pfam: PF00377; Prion; 1.

DR Pfam: PF03991; Prion octapep; 5.

DR PRINTS: PR00341; PRION.

DR SMART: SM00086; PAC; 1.

DR SMART: SM00157; PRP; 1.

DR PROSITE: PS00291; PRION 1; 1.

DR PROSITE: PS00706; PRION 2; 1.

KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.

FT SIGNAL 1 24

FT CHAIN 25 256 MAJOR PRION PROTEIN 2. (POTENTIAL).

FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFID 182 217 BY SIMILARITY.

FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-O.

FT REPEAT 54 62 0.

FT REPEAT 63 70 1.

FT REPEAT 71 78 2.

FT REPEAT 79 86 3.

FT REPEAT 87 95 4.

FT REPEAT 87 95 5.

SO SEQUENCE 256 AA; 27880 MW; 0D969F2D9033B30 CRC64;

Query Match 8.94; Score 93.5; DB 1; Length 256;

Best Local Similarity 29.24; Pred. No. 0.5;

Matches 47; Conservative 6; Mismatches 57; Indels 51; Gaps 12;

QY 17 RREPTDKPFGGGQIVGGVYLLPRGRLGVATRTKTSERSQPRGRQPIPKA---RQFQ 72

DB 26 KRPK---PGGWNMG-----SRYPGG-----SPGGRNRP-FQGGGWCQPH 64

QY 73 GRHMADP---GYWPLXNMGCGW-----AGWLSPRGSRPRHWGNDPRRRRNIGKV 123

DB 65 GGGWGGPHGGGWCQWQ---HGGWGGPHGGGWCW---GQGGSHSQW---NKSPKPTNMKQV 116

QY 124 DTLTCGFADLMGYIPVVGAPLGGVAALAHGVAIEDGINY 164

DB 117 GAAAG-----AVVGG-LGGYMLGSAMSRPLIFGNDY 148

RESULT 23

PRIO_BOVIN STANDARD; PRT; 264 AA.

ID PRIO_BOVIN STANDARD; PRT; 264 AA.

AC P10279;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Major prion protein 1 precursor (Prp) (Major scrapie-associated fibril protein 1).

DE PNP OR PRP.

GN PNP OR PRP.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

OC NCBI_TaxID=9913;

OX NCBI_TaxID=9913;

RA MEDLINE=93118243; PubMed=1362024;

RA Yoshimoto J., Iinuma T., Ishiguro N., Horiuchi M., Imanura M., Shingawa M.;

RA "Comparative sequence analysis and expression of bovine Prp gene in mouse L-929 cells."

RL Virus Genes 6:343-356(1992).

CC -1- FUNCTION: The function of Prp is not known. Prp is encoded in the host genome and is expressed both in normal and infected cells.

CC -1- SUBUNIT: Prp has a tendency to aggregate yielding polymers called "rods".

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.

CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU, CREDITFIELD-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE), TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.

CC -1- SIMILARITY: Belongs to the prion family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: J01614; BAA01469.1; -

DR PIR: J02668; J02668.

DR HSSP: P10279; JDMY.

DR InterPro: IPR001610; PAC.

DR InterPro: IPR000817; Prion.

DR Pfam: PF00377; Prion; 1.

DR Pfam: PF03991; Prion octapep; 5.

DR PRINTS: PR00341; PRION.

DR SMART: SM00086; PAC; 1.

DR SMART: SM00157; PRP; 1.

DR PROSITE: PS00291; PRION 1; 1.

DR PROSITE: PS00706; PRION 2; 1.

KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.

FT SIGNAL 1 24

FT CHAIN 25 256 MAJOR PRION PROTEIN 2. (POTENTIAL).

FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFID 182 217 BY SIMILARITY.

FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-O.

FT REPEAT 54 62 0.

FT REPEAT 63 70 1.

FT REPEAT 71 78 2.

FT REPEAT 79 86 3.

FT REPEAT 87 95 4.

FT REPEAT 87 95 5.

SO SEQUENCE 256 AA; 27880 MW; 0D969F2D9033B30 CRC64;

Query Match 8.94; Score 93.5; DB 1; Length 256;

Best Local Similarity 29.24; Pred. No. 0.5;

Matches 47; Conservative 6; Mismatches 57; Indels 51; Gaps 12;

RN [7]
 RP SEQUENCE OF 1-15 FROM N.A.
 RA Tanaka M., Inoue S., Ikeda T., Horiuchi M., Ishiguro N., Shinagawa M.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 25-36.
 RA MEDLINE=89057122; PubMed=2904126;
 RA Hope J., Reekie L.J.D., Hunter N., Multchaup G., Beyreuther K.,
 RA White H., Scott A.C., Stack M.J., Dawson M., Wells G.A.;
 RT "Fibrils from brains of cows with new cattle disease contain scrapie-
 associated protein.";
 RL Nature 336:390-392(1988).
 RN [9]
 RP STRUCTURE BY NMR OF 132-241.
 RX MEDLINE=20359707; PubMed=10899999;
 RA Lopez Garcia F., Zahn R., Riek R., Muechtlich K.;
 RL "NMR structure of the bovine prion protein.";
 RT Proc. Natl. Acad. Sci. U.S.A. 97:8334-8339(2000).
 CC -1- FUNCTION: The function of PrP is not known. PrP is encoded in the
 CC host genome and is expressed both in normal and infected cells.
 CC -1- SUBUNIT: PrP has a tendency to aggregate yielding polymers called
 CC "rods".
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS
 CC INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,
 CC CREATZFELDT-JAKOB DISEASE (CJD), GERSTMAN-STAUSSELER SYNDROME
 CC (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
 CC TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -1- SIMILARITY: Belongs to the prion family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55882; CAA39368.1; -;
 DR EMBL; D10612; BAA01467.1; -;
 DR EMBL; D10613; BAA01468.1; -;
 DR EMBL; S55629; AAB25514.1; -;
 DR EMBL; AB001468; BAA19253.1; -;
 DR EMBL; AJ298878; CAC37367.1; -;
 DR EMBL; AF517842; AAM66709.1; -;
 DR EMBL; D26151; BAA05138.1; -;
 DR PIR; A54330; A54330.
 DR PDB; 1DWY; 26-FEB-02.
 DR PDB; 1DWZ; 26-FEB-02.
 DR PDB; 1DX0; 26-FEB-02.
 DR PDB; 1DX1; 26-FEB-02.
 DR InterPro: IPR000817; Prion.
 DR Pfam: PF00377; prion; 1.
 DR Pfam: PF03991; prion_octapep; 6.
 DR PRINTS; PR00341; PRION.
 DR SMART; SM00157; PRP; 1.
 DR PROSITE; PS00291; PRION_1; 1.
 DR PROSITE; PS00706; PRION_2; 1.
 KM Prion; Glycoprotein; GPI-anchor; Repeat; Signal; Polymorphism;
 KW 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 264 MAJOR PRION PROTEIN 1.
 FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT DISULFID 190 225
 FT DOMAIN 54 103
 FT REPEAT 54 62
 FT REPEAT 63 70
 FT REPEAT 71 78
 FT REPEAT 79 86
 FT REPEAT 87 94
 FT REPEAT 95 103
 FT REPEAT 95 103

FT VARIANT 71 78 MISSING (IN ALLELE 2).
 FT CONFLICT 218 218 E -> K (IN REF. 2 AND 4).
 FT HELIX 136 138
 FT STRAND 140 141
 FT HELIX 155 162
 FT TURN 163 164
 FT HELIX 165 167
 FT STRAND 173 174
 FT HELIX 184 203
 FT TURN 204 206
 FT HELIX 211 237
 SQ SEQUENCE 264 AA; 28614 MW; D6D214038316A231 CRC64;
 Query Match 8.9%; Score 93.5; DB 1; Length 264;
 Best local Similarity 28.1%; Pred. No. 0.51;
 Matches 48; Conservative 7; Mismatches 53; Indels 63; Gaps 13;
 QY 17 RRPDYKFPFGGQYVGVYLLRRRGRGLGVRATRKTSRSGRRRQPIPKA---RQIQ 72
 DB 26 KRPK---PGGWMNTGG-----SRYPGQG-----SPQGNRYP-PQGGGQWGP 64
 QY 73 GRHMGP---GYWPVLYNMGCGW-----AGWLLSPRGRPH---WGP-----NDPR 113
 DB 65 GGGWGPFGGQWGP---HGCGWGPFGCGW-----GQPHGGGQWGGGTHGQNNKPS 114
 QY 114 RSRNLKGYIDLTGCFADLMGYPVGPAPLGVAALAGVRALEDGINY 164
 DB 115 KPTTNKHVAGAAAG-----AVVGG-LGGYMGSMNSRPLHFSGDY 156
 RESULT 24
 PRP1 TRAST
 ID PRP1 TRAST STANDARD; PRT; 264 AA.
 AC P40242;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Major prion protein 1 precursor (PrP) (Major scrapie-associated fibril
 DE protein 1).
 OS Tricelaphus strepsiceros (Greater kudu).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Tricelaphus.
 OX NCBI_Taxid=9946;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Martin T.C., Hughes S.L., Hughes K.J., Dawson M.;
 RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: The function of PrP is not known. PrP is encoded in the
 CC host genome and is expressed both in normal and infected cells.
 CC -1- SUBUNIT: PrP has a tendency to aggregate yielding polymers called
 CC "rods".
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS
 CC INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,
 CC CREATZFELDT-JAKOB DISEASE (CJD), GERSTMAN-STAUSSELER SYNDROME
 CC (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
 CC TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -1- SIMILARITY: Belongs to the prion family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74771; CAA52761.1; -;
 DR PIR; S37137; S37137.
 DR HSSP; P10279; 1DWY.
 DR InterPro: IPR000817; Prion.

DR Pfam; PF00377; prion; 1.
 DR Pfam; PF03991; prion octapep; 6.
 DR PRINTS; PR00341; PRION.
 DR SMART; SM00157; PRP; 1.
 DR PROSITE; PS00291; PRION_1; 1.
 DR PROSITE; PS00706; PRION_2; 1.
 KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 264 MAJOR PRION PROTEIN 1. (POTENTIAL).
 FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 190 225 BY SIMILARITY.
 FT DOMAIN 54 103 6 X 8 AA TANDEM REPEATS OF P-H-G-G-W-G-
 FT REPEAT 54 62 0.
 FT REPEAT 63 70 1.
 FT REPEAT 71 78 2.
 FT REPEAT 79 86 3.
 FT REPEAT 87 94 4.
 FT REPEAT 95 103 5.
 FT REPEAT 95 103 6.
 SQ SEQUENCE 264 AA; 28644 MW; FEB73F4173261991 CRC64;
 Query Match 8.8%; Score 91.5; DB 1; Length 264;
 Best Local Similarity 27.6%; Pred. No. 0.76;
 Matches 47; Conservative 6; Mismatches 56; Indels 61; Gaps 12;
 QY 17 RRPDVVFPGGGQIVGVVLLPRGPRLVATRKTSERSQPRGRPIPKA---RPOG 73
 DB 26 KRPR---PGGWMNTG---SRYPGQG---SPGGRNRYPSGGGGWGP 65
 QY 74 RHNAQP---GYPWPLYNGECGW---AGWLLSPRGRPH---WGP---NDPRR 114
 DB 66 GCGMGQPHGGMGQP---HGGMGQPHGGW---GQHGGGGWGGGTHGGMNPK 115
 QY 115 RSNRLKGVDTLTGCFADLMGIYVVGAPLVGVAALAHGVRALDGINY 164
 DB 116 PRTNMKIVAGAAAG---AVVGG-LGVVLSGMSRPLIHFSYD 156
 RESULT 25
 PRIO_RABIT STANDARD; PRT; 252 AA.
 AC 095211;
 DT 01-NOV-1997 (Rel. 35, Last Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Major prion protein precursor (PrP) (PrP27-30) (PrP33-35C).
 PRNP OR PRP.
 OC Oryzctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand white;
 RX MEDLINE=97183665; PubMed=9031631;
 RA Loftus B., Rogers M.;
 RT "Characterization of a prion protein (PrP) gene from rabbit: a
 species with apparent resistance to infection by prions.";
 RL Gene 184;215-219(1997).
 CC -1- FUNCTION: The function of PrP is not known. PrP is encoded in the
 host genome and is expressed both in normal and infected cells.
 CC -1- SUBUNIT: PrP has a tendency to aggregate yielding polymers called
 "rds".
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- DISEASE: PrP is FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND
 ANIMALS INFECTED WITH THE DEGENERATIVE NEUROLOGICAL DISEASES KURU,
 CREUTZFELDT-JAKOB DISEASE (CJD), GERSTMAN-STRAUSSLER SYNDROME
 (GSS), SCRAPE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
 CC TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -1- SIMILARITY: Belongs to the prion family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; U2834; AAC48697.1; -.
 DR PIR; JC6175; JC6175.
 DR HSSP; P10279; IDWY.
 DR InterPro; IPR000817; Prion.
 DR Pfam; PF00377; prion; 1.
 DR Pfam; PF03991; prion octapep; 5.
 DR PRINTS; PR00341; PRION.
 DR SMART; SM00157; PRP; 1.
 DR PROSITE; PS00291; PRION_1; 1.
 DR PROSITE; PS00706; PRION_2; 1.
 KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 252 MAJOR PRION PROTEIN.
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 178 213 BY SIMILARITY.
 FT DOMAIN 51 92 5 X 8 AA TANDEM REPEATS OF P-H-G-G-W-G-
 FT REPEAT 51 59 0.
 FT REPEAT 60 67 1.
 FT REPEAT 68 75 2.
 FT REPEAT 76 83 3.
 FT REPEAT 84 92 4.
 FT REPEAT 84 92 5.
 SQ SEQUENCE 252 AA; 27432 MW; 2E177A4F38B23A54 CRC64;
 Query Match 8.7%; Score 91; DB 1; Length 252;
 Best Local Similarity 26.4%; Pred. No. 0.8;
 Matches 38; Conservative 12; Mismatches 38; Indels 56; Gaps 10;
 QY 17 RRPDVVFPGGGQIVGVVLLPRGPRLVATRKTSERSQPRGRPIPKARPOGRHW 76
 DB 24 KRPR---FOGGWMNTG---SRYPGQG---SPGGRNRYPSGGGGWGP 57
 QY 77 AQP---GYPWPLYNGN---EGCGW---AGWLLSPRGRPHWG---NDPRRSRLNIGKVI 123
 DB 58 GQPHGGMGQPHGGMGQPHGGMGQPHGGW---GQGTINQWKGKPSKPTSMKHV---- 111
 QY 124 DTLTGFADLMGIYVVGAPLVGV 147
 DB 112 ---AGAA-----AAGAVVGG 124
 RESULT 26
 PRIO_PIG STANDARD; PRT; 257 AA.
 AC P49927;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Major prion protein precursor (PrP).
 PRNP.
 GN Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95244627; PubMed=7727546;
 RA Martin T., Hughes S., Hughes K., Dawson M.;
 RT "Direct sequencing of PCR amplified pig PrP genes.";
 RL Biochim. Biophys. Acta 1270:211-214(1995).
 CC -1- FUNCTION: The function of PrP is not known. PrP is encoded in the
 host genome and is expressed both in normal and infected cells.
 CC -1- SUBUNIT: PrP has a tendency to aggregate yielding polymers called
 "rds".
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

Db 110 TNNKHVAGAAAG-----AVVGG-IGGYMLGSAMSRPLIHFGNDY 148

RESULT 28

PRT: 256 AA.

AC P47852; STANDARD; PRT; 256 AA.

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Major prion protein precursor (PrP).

GN PRNP.

OS Odocoileus hemionus (Mule deer) (Black-tailed deer).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae; Cervidae; Odocoileinae; Odocoileus.

OX NCBI_TaxID=9872;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=97393774; PubMed=9250209;

RT Cervenkova L., Rohwer R., Williams S., Brown P., Gajdusek D.C.;

RT "High sequence homology of the PrP gene in mule deer and Rocky Mountain elk."

RL Lancet 350:219-220(1997).

CC -1- FUNCTION: The function of PrP is not known. PrP is encoded in the host genome and is expressed both in normal and infected cells.

CC -1- SUBUNIT: PrP has a tendency to aggregate yielding polymers called "yods".

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By similarity).

CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF ANIMALS INFECTED WITH THE DEGENERATIVE NEUROLOGICAL DISEASE KNOWN AS CHRONIC WASTING DISEASE (CMD).

CC -1- SIMILARITY: Belongs to the prion family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U25965; AAA68941.1; -.

DR HSSP; P10279; IDWY.

DR InterPro; IPR000817; Prion.

DR Pfam; PF00377; Prion; 1.

DR Pfam; PF03991; Prion octapep; 5.

DR PRINTS; PR00341; PRION.

DR SMART; SM00157; PRP; 1.

DR PROSITE; PS00291; PRION_1; 1.

DR PROSITE; PS00706; PRION_2; 1.

KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.

FT SIGNAL 1 24 BY SIMILARITY.

FT CHAIN 25 256 MAJOR PRION PROTEIN.

FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFID 182 217 BY SIMILARITY.

FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-O.

FT REPEAT 54 62 1. POTENTIAL.

FT REPEAT 63 70 2. MAJOR PRION PROTEIN 2.

FT REPEAT 71 78 3. N-LINKED (GLCNAC. . .) (POTENTIAL).

FT REPEAT 79 86 4. N-LINKED (GLCNAC. . .) (POTENTIAL).

FT REPEAT 87 95 5. BY SIMILARITY.

FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-O.

SO SEQUENCE 256 AA; 27961 MW; B98EB121C302FD36 CRC64;

Query Match 8.7%; Score 90.5; DB 1; Length 256;

Best Local Similarity 27.4%; Pred. No. 0.89; 51; Indels 65; Gaps 12;

Matches 46; Conservative 6; Mismatches 51; Indels 65; Gaps 12;

Qy 17 RRPDVPFGGGQIVGVYLLPRRGPRIGVYATKTSERSQPRGRPIPKA-----RQPG 72

Db 26 KRPK-----PGGWNITGG-----SRYPQG-----SPGNNRYP-PGGGGWGQPH 64

Qy 73 GRHMAQPGYWPPLXNHEGCGW-----AGWLLSRGRNPH-----WGP-----NDPERRS 116

Db 65 GGGWGQP-----HGGGQPHGGW-----QQPHGGGQGGGTHSQNNKPSKPK 109

Qy 117 RNIGKVIDTLTCGFADLMGYIPVVGAPILGGAALAHGVAIEDGINY 164

Db 110 TNNKHVAGAAAG-----AVVGG-IGGYMLGSAMSRPLIHFGNDY 148

RESULT 29

PRT: 256 AA.

AC P40243; STANDARD; PRT; 256 AA.

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Major prion protein 2 precursor (PrP) (Major scrapie-associated fibril protein 2).

OS Tragelaphus strepsiceros (Greater kudu).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Tragelaphus.

OX NCBI_TaxID=9946;

RN [1]

RP SEQUENCE FROM N.A.

RA TISSUE=Brain;

RA Martin T.C., Hughes S.J., Dawson M.;

RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: The function of PrP is not known. PrP is encoded in the host genome and is expressed both in normal and infected cells.

CC -1- SUBUNIT: PrP has a tendency to aggregate yielding polymers called "yods".

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.

CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU, CREUTZFELDT-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE), TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.

CC -1- SIMILARITY: Belongs to the prion family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X74759; CA52775.1; -.

DR HSSP; P10279; IDWY.

DR InterPro; IPR000817; Prion.

DR Pfam; PF00377; Prion; 1.

DR Pfam; PF03991; Prion octapep; 5.

DR PRINTS; PR00341; PRION.

DR SMART; SM00157; PRP; 1.

DR PROSITE; PS00291; PRION_1; 1.

DR PROSITE; PS00706; PRION_2; 1.

KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.

FT SIGNAL 1 24 BY SIMILARITY.

FT CHAIN 25 256 MAJOR PRION PROTEIN 2.

FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFID 182 217 BY SIMILARITY.

FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-O.

FT REPEAT 54 62 1. POTENTIAL.

FT REPEAT 63 70 2. MAJOR PRION PROTEIN 2.

FT REPEAT 71 78 3. N-LINKED (GLCNAC. . .) (POTENTIAL).

FT REPEAT 79 86 4. N-LINKED (GLCNAC. . .) (POTENTIAL).

FT REPEAT 87 95 5. BY SIMILARITY.

FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-O.

SO SEQUENCE 256 AA; 28050 MW; D4D02CDBFC918743 CRC64;

